12-9-2006

Molecular-Genetic and Structural Analyses of the NifHDKX Proteins of the Nitrogenase System

Surobhi Lahiri

MOLECULAR-GENETIC AND STRUCTURAL ANALYSES OF THE NIFHDKX PROTEINS OF THE NITROGENASE SYSTEM

By

Surobhi Lahiri

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Biology
in the Department of Biological Sciences

Mississippi State, Mississippi

December 2006

MOLECULAR-GENETIC AND STRUCTURAL ANALYSES OF THE NIFHDKX PROTEINS OF THE NITROGENASE SYSTEM

By

Surobhi Lahiri

Approved:

Narasaiah Gavini Professor of Biological Sciences (Advisor, Director of Dissertation)	Dwayne Wise Professor of Biological Sciences (Committee Member, Graduate Coordinator)
John Boyle Professor of Molecular Biology Biology and Biochemistry (Committee member)	Lakshmidevi Pulakat Professor of Biological Sciences (Committee member)
Donald Downer Professor of Biological Sciences (Committee Member)	Philip B. Oldham Dean of the College of Arts and Sciences

Name: Surobhi Lahiri

Date of Degree: December 09, 2006

Institution: Mississippi State University

Major Field: Biological Sciences

Major Professor: Dr. Nara Gavini

Title of Study: MOLECULAR-GENETIC AND STRUCTURAL ANALYSES OF THE NIFHDKX PROTEINS OF THE NITROGENASE ENZYME

Pages in Study: 208

Candidate for Degree of Doctor of Biology

The nitrogenase enzyme is the biochemical machinery responsible for the conversion of the largely unavailable nitrogen to the easily assimilable ammonia for living organisms by the process termed as biological nitrogen fixation (BNF). This study was focused on understanding the various structural and functional aspects of the nitrogenase enzyme related to maturation and assembly of the FeMo-cofactor (FeMoco) metallocluster of the MoFe protein (the site for final substrate reduction), development of a dimeric MoFe protein and the structural homology of nitrogenase with other metalloenzymes. This research was specifically directed towards the NifHDKX proteins in which the *nifHDK* genes are the major structural genes that encode the nitrogenase enzyme and *nifX* is an accessory gene that encodes the NifX protein, indicated to be involved in the formation of the FeMoco. The overall objective of this study was to gain

structural and functional information on the nitrogenase enzyme through the study of the NifHDKX proteins. A major part of our study included the detection of protein-protein interactions between the NifD, NifK and a fused NifDK protein. The results of this study could prove to be useful for further studies that are directed towards condensing the *nif* genes so as to facilitate transfer of nitrogen fixing genes to plants for their improved nutrition. We also determined protein-protein interactions between NifX and other proteins involved in the FeMoco biosynthetic pathway. Based on the results, we were able to describe the role of NifX and propose a modified model for the FeMoco biosynthesis pathway. Apart from this, a comparative structural and evolutionary study was performed on the NifH similar proteins such as ChlL, CompA, MinD and ArsA and the NifDK similar proteins known as ChlBN. Based on the conservation of similar structural domains in NifH and ArsA, NifH was found to complement the function of ArsA1. Also the comparison between NifDK and the homology modeled ChlBN protein structure suggested a potential site for the presence of a FeMoco in ChlN. Thus, these studies helped us to derive meaningful conclusions on the structure and evolution of the nitrogenase enzyme and its homologs in nature.

DEDICATION

Dedicated to my grandparents:

Late Dr. B. B. Lahiri, Late Mrs. Urmila Lahiri
Late Mr. M. M. Chakraborty, Mrs. Geeta Chakraborty

'Your Dreams Are My Motivation'

ACKNOWLEDGMENTS

First and foremost I would like to express my gratefulness to my advisor, Dr. Nara Gavini, for his valuable guidance and continuous motivation. Throughout my PhD studies, he has emphasized on the importance of a scientific and professional attitude, hard work and independent thinking. He will always remain an inspiration for all endeavors in my career. I take this opportunity to thank him and hope that this dissertation measures upto his expectations.

I am indebted to Dr. Lakshmidevi Pulakat (co-advisor, committee member) for her supervision of my projects during the entire course of my Ph.D. She has shown by her own example the meaning of will-power, perseverance and dedication to work. I thank you for your direction and your valuable time spent in editing my manuscripts.

I am also very fortunate to have had the most well-distinguished professors from various scientific areas as my committee members. They have not only instilled in me further interest for science but also encouraged and guided me to aim for perfection. I am therefore indeed very grateful to Dr. Donald Downer, Dr. John Boyle and Dr. Dwayne Wise.

Ms. Gloria Blankenship (graduate secretary, Biology Dept., MSU) and Ms. Loraine DeVenney (graduate secretary, Biology Dept., BGSU) deserve my special thanks

for all their help in official matters and for always putting a smile on my face. I also thank JoAnne, Mable and Tom for all their help.

I would like to heartily acknowledge the contribution of my friends and lab colleagues who have made such a positive difference in strongly pursuing my Ph.D. I would most of all like to thank Vandana and Preeti who have remained wonderful friends throughout my studies in BGSU and later here in MSU. I also thank Kumar, Han, Lakshmi, Girish and Swapna for their friendship and support.

My brother Ronny has lent me the lighter moments during the past few years. Also he has made me aim higher because I knew that he was looking up to me for inspiration. I thank him for his affection and his faith in me. Words cannot suffice the extent of support and encouragement that I have received from my parents-in-law and my brother and sister-in-law. Mummyji, Papaji, Dada and Bhabhi -- I will always remember the confidence you showed in me and the way you made my path to success easier.

Baba and Maa -- can I even begin to thank you? You have been my pillar of strength in life throughout and I hope this accomplishment will make you proud. With love and respect, this work is my gift to you …..

Finally, the person who has been the strongest basis for my continuing to strive for this degree is my husband, Vikas. Not only has he sacrificed two years of his life staying apart while I was completing my Ph.D. here but he has also traveled this journey with me as a friend and guide. Thank you, Vikas…

And then again, I thank you, Lord Shri Krishna, for this day…

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Nitrogen (N) is present in many chemical forms, both organic and inorganic, in the atmosphere, biosphere, hydrosphere, and geosphere. The approximate global inventory of N in the four spheres is shown in Table 1.1. N occurs in the gas, liquid (dissolved in water), and solid phases. It can be associated with carbon (organic species) and with elements other than carbon (inorganic species). Important inorganic species include dinitrogen (N_2), nitric acid (HNO_3), nitrate (NO_3^-), nitrite (NO_2^-), nitrous oxide (N_2O), nitric oxide (NO), N dioxide (NO_2), ammonium (NH_4^+), and ammonia (NH_3). Most organic N species in the four spheres are biomolecules, such as proteins, peptides, enzymes, and genetic material (RNA and DNA) (Furley and Newey, 1983).

Table 1.1. Approximate global inventory of N in the four spheres

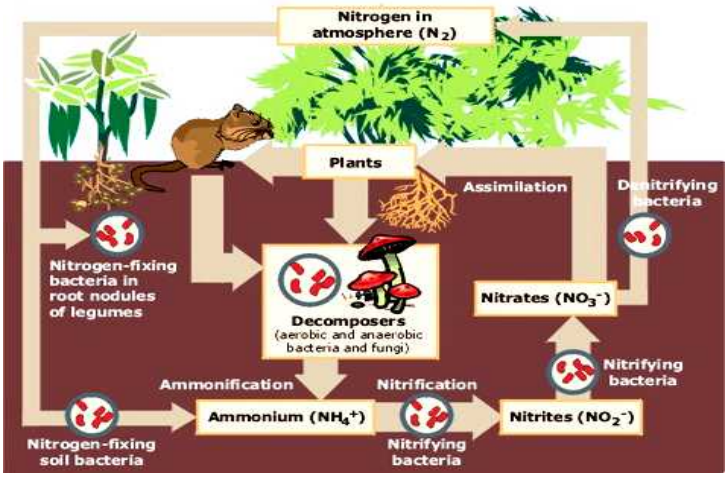
Sizes of Global N Reservoir

Reservoir/Pool Type	**Metric Tons**	**% of Total**
Biosphere	2.8×10^{11}	0.0002
Hydrosphere	2.3×10^{13}	0.014
Atmosphere	3.86×10^{15}	2.3
Geosphere	1.636×10^{17}	97.7
Crust	$0.13 - 1.4 \times 10^{16}$	0.78-8.4
Soils and Sediments	$0.35 - 4.0 \times 10^{15}$	0.21-2.4
Mantle and Core	1.6×10^{17}	95.6

The bulk of the N (about 98%) exists in the geosphere, and most of the remainder is found in the atmosphere. The N in the biosphere is highly reactive and rapidly cycled (http://www.sws.uiuc.edu/nitro/nitrodesc.asp)

The Nitrogen cycle

Five main processes cycle nitrogen through the biosphere, atmosphere, and geosphere: nitrogen fixation, nitrogen uptake (organismal growth), nitrogen mineralization (decay), nitrification and denitrification (Stevenson, 1986). Microorganisms, particularly bacteria, play major roles in all of the principal nitrogen transformations. The nitrogen transformation rates are affected by environmental factors that influence microbial activity, such as temperature, moisture, and resource availability. The nitrogen cycle is thus described as the continuous flow of nitrogen through the biosphere by the processes of nitrogen fixation, ammonification (decay), nitrification, and denitrification (Fig. 1.1). Human activities also influence the N cycle through interacting physical, chemical and biological processes (Stevenson, 1982).



Schematic representation of the flow of nitrogen through the environment. The importance of bacteria in the cycle is immediately recognized as being a key element in the cycle, providing different forms of nitrogen compounds assimilable by higher organisms (http://www.epa.gov/maia/html/nitrogen.html)

Fig. 1.1 The nitrogen cycle.

The major human activities that today influence the global N cycle are fossil fuel combustion, the production and use of chemical fertilizer, and the growing of N-fixing crops. These activities are reported to have doubled the magnitude of N fixation over continents (Galloway et. al., 1995, Chapin et al., 2002).

Nitrogen fixation

In order for nitrogen to be used for growth it must be "fixed" (combined) in the form of ammonium (NH_4^+) or nitrate (NO_3^-) ions. Nitrogen fixation is the process in which atmospheric N_2 is converted to compounds that can be utilized by plants. It is both a natural process that is mediated by certain microorganisms and an industrial process that requires large amounts of energy. Although all organisms require N compounds, very few are able to utilize N_2, the most abundant, readily-available form of the element. Most organisms require fixed forms of N in the form of NH_3, NO_3^-, NO_2^-, or organic-N. So, nitrogen is often the limiting factor for growth and biomass production in all environments where there is suitable climate and availability of water to support life (Pratt, 1978).

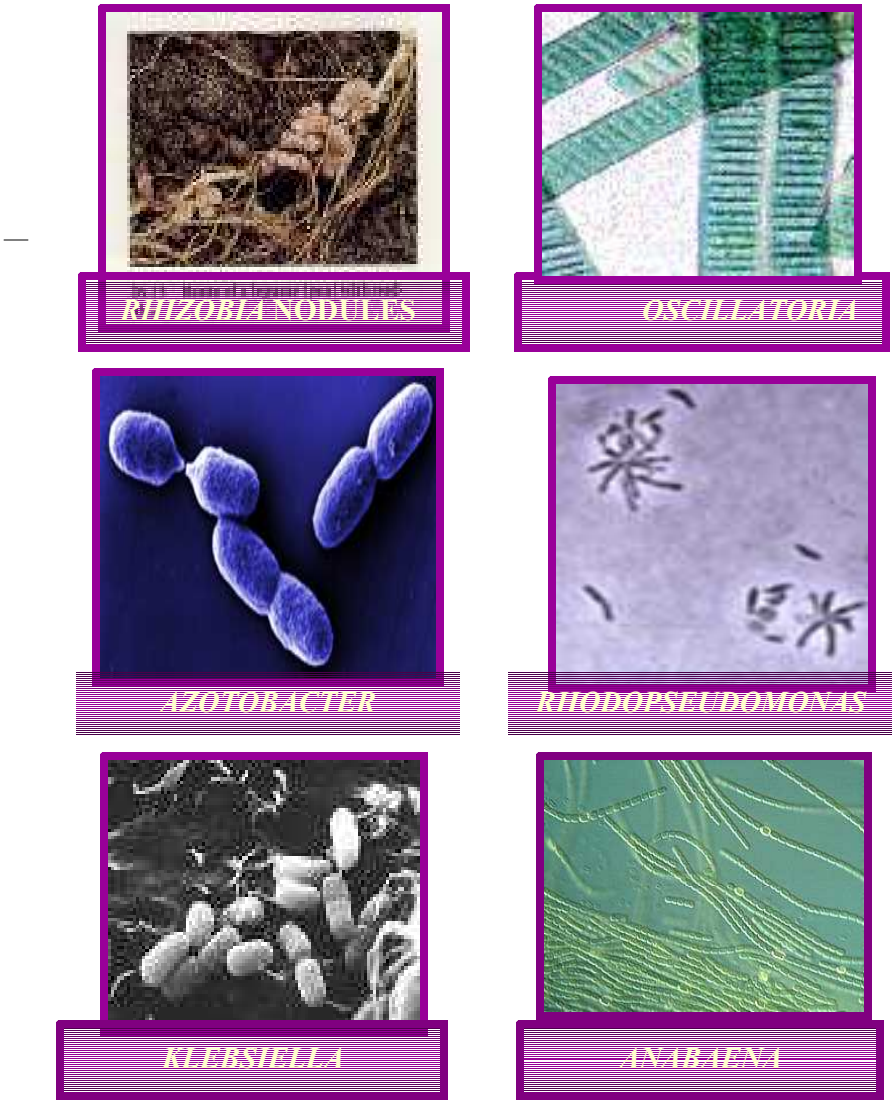
Biological nitrogen fixation (BNF) provides the dominant route for the transformation of atmospheric dinitrogen into a bioavailable form, ammonia. Fixed nitrogen in the form of ammonia (NH_3) is a critical starting material for fundamental processes for both agriculture and industry. BNF, effected primarily by diazotrophic bacteria, supplies more than 60% of the world's annual resources of new ammonia (Schlesinger, 1991). When compared to the commercial chemical Haber-Bosch process reaction with its high

temperature and pressure requirements, the biological process of producing ammonia from dinitrogen presents a striking contrast. Nitrogen fixation has also been achieved by using a non-metallic buckminsterfullerene (C_{60}) molecule, in the form of a water-soluble C_{60}:γ -cyclodextrin (1:2) complex, and light under nitrogen at atmospheric pressure. This metal-free system efficiently fixed nitrogen under mild conditions by making use of the redox properties of the fullerene derivative (Nishibayashi et al., 2004).

Modern agriculture relies heavily on chemical fertilizers. Synthetic nitrogen use has increased 10-fold over the last 40 years, representing worldwide cost of over $20 billion. Industrial production of nitrogen fertilizer consumes large amounts of natural gas and releases carbon dioxide, which probably hastens global warming. Fertilizers also drain into ground water, streams, and rivers, thus risking human and animal health. BNF and nitrogen-fixing bacteria can play critical roles in achieving environmentally safe and economical farming systems. Exploiting the biochemistry and genetics of nitrogen-fixing bacteria helps to lessen our dependence on nitrogen fertilizers.

Nitrogen-fixing bacteria (diazotrophs) are important in the maintenance of the biogeochemical nitrogen cycle. All diazotrophs identified so far are prokaryotes distributed widely in the archaeal and bacterial domains (Fig. 1.2). Well-characterized nitrogen-fixing systems have been found in some free-living species of cyanobacteria (e.g. *Trichodesmium*), methanogens (e.g. *Methanococcus*), obligate aerobes (e.g. *Azotobacter*), facultative anaerobes (e.g. *Klebsiella*), and obligate anaerobes (e.g. *Clostridium*). Diazotrophs may also reside in symbiotic relationships with plants or (rarely) lichens. Nitrogen fixing bacteria belonging to genera such as *Rhizobium*,

Bradyrhizobium, etc. are commonly found in root nodules on plants (mostly legumes). Nitrogen- fixing bacteria (or diazotrophs) play a central role in almost all aspects of nitrogen availability and thus for life support on earth (Postgate, 1998).



Images from 'http://images. google.com/images?q=nitrogen+fixing+bacteria'

Fig. 1.2. Examples of free-living or symbiotic diazotrophic bacteria.

Nitrogen uptake

The ammonia produced by nitrogen fixing bacteria is usually quickly incorporated into protein and other organic nitrogen compounds, either by a host plant, the bacteria itself, or another soil organism (Rudolf and Kroneck, 2005 (*review*)).

Nitrogen mineralization

After nitrogen is incorporated into organic matter, it is often converted back into inorganic nitrogen by a process called nitrogen mineralization, otherwise known as decay. When organisms die, decomposers (such as bacteria and fungi) consume the organic matter and lead to the process of decomposition. During this process, a significant amount of the nitrogen contained within the dead organism is converted to ammonium

Nitrification

Some of the ammonium produced by decomposition is converted to nitrate via a process called nitrification. The bacteria that carry out this reaction gain energy from it (eg. *Nitrosomonas* and *Nitrobacter*) (Rudolf and Kroneck, 2005 (*review*)).

Denitrification

Through denitrification, oxidized forms of nitrogen such as nitrate and nitrite (NO_2^-) are converted to dinitrogen (N_2) and, to a lesser extent, nitrous oxide gas. Denitrification is an anaerobic process that is carried out by denitrifying bacteria (eg.

Thiobacillus denitrificans, *Micrococcus denitrificans, Pseudomonas sp.*) which convert nitrate to dinitrogen in the following sequence:

$NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$.

Once converted to dinitrogen, nitrogen is unlikely to be reconverted to a biologically available form because it is a gas and is rapidly lost to the atmosphere. Denitrification is the only nitrogen transformation that removes nitrogen from ecosystems and it roughly balances the amount of nitrogen fixed by diazotrophs (http://helios.bto.ed.ac.uk/bto/microbes/nitrogen.htm)

The nitrogenase enzyme

The enzyme responsible for the conversion of dinitrogen to ammonia in nitrogen-fixing organisms is known as nitrogenase (E. C. 1.18.6.1). Dinitrogen is relatively inert due to its triple bond. The enzyme therefore requires a lot of chemical energy in the form of ATP and reducing agents, such as NADH. The enzyme is composed of two components, known as the Iron (Fe) and Iron-Molybdenum (MoFe) proteins. The MoFe protein is supplied reducing power upon its association with the reduced, nucleotide-bound Fe protein. The nitrogenase complex undergoes steps of association and disassociation to transfer one electron, which is the limiting step in the process. Three such electron transfers are necessary for one complete cycle of nitrogen fixation. The process itself is very energy-intensive, requiring electron donors and ATP to provide reducing power. The final nitrogen reduction reaction is thus described as given below:

$$N_2 + 8H^+ + 8e^- + 16\,ATP \rightarrow 2\,NH_3 + H_2 + 16ADP + 16\,Pi$$

Nitrogenase is able to bind acetylene and carbon monoxide, which are noncompetitive substrates and inhibitors, respectively. Dinitrogen, however, is a competitive substrate for acetylene (Seefeldt et. al., 2004).

Genetic, structural, mechanistic and functional aspects of nitrogenase

Although nitrogen fixation is a property of a phylogenetically diverse set of bacteria and cyanobacteria, in general, the sequences, structures, and functional properties of the nitrogenase are highly conserved between different organisms (Rees, 1993; Peters et al., 1995; Howard and Rees, 1996). Four classes of nitrogenases have been characterized to date (Stiefel, 1996; Hofmann-Findeklee, 2000).

1. Molybdenum Nitrogenases – The Mo-dependent nitrogenases are the best studied and most widely distributed. This enzyme consists of two oxygen sensitive metalloproteins, a Fe protein containing a single [4Fe-4S] center, which acts as an obligate ATP dependent electron donor to the MoFe protein. The MoFe protein has two types of redox centers, two P clusters consisting of two bridged [4Fe-4S] cubanes, and two Fe- and Mo-containing cofactor centers (FeMoco), the probable site of nitrogen binding and reduction (Christiansen et al., 2001)
2. Vanadium Nitrogenases – A genetically distinct nitrogenase system in which the conventional molybdoprotein was replaced by a vanadoprotein was discovered in *Azotobacter chroococcum* and *Azotobacter vinelandii*. Both Mo-nitrogenases and V-nitrogenases were found to have similar requirements for activity: MgATP, a low

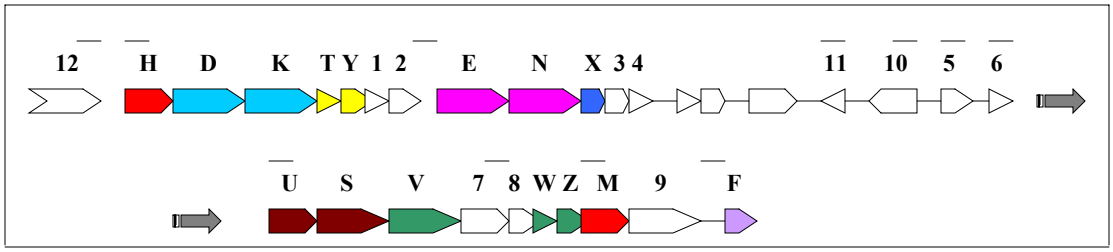
potential reductant and the absence of oxygen. The genes encoding the V-nitrogenase are expressed only under conditions of Mo-deficiency (Eady, 1988).

3. Iron Nitrogenases – The third alternative form of nitrogenase in *Azotobacter vinelandii* is synthesized under conditions of molybdenum and vanadium deficiency and known as an iron only nitrogenase. Studies have shown that the only transition metal in the iron nitrogenase is iron and that it is the only minimal metal requirement for its synthesis (Eady, 1996; Chisnell et al., 1988; Pau et al., 1989).
4. Superoxide dependent Nitrogenase - Ribbe *et al.* investigated yet another, rare nitrogenase system called the superoxide dependent nitrogenase that exhibited features which were not obvious in the diazotrophic bacteria studied till then (Ribbe et al., 1997). This nitrogenase found only in *Streptomyces thermoautotrophicus*, contains a Mo-molybdopterin cytosine dinucleotide cofactor (Mo-MCD) as the active site and fixes dinitrogen using CO or H_2 and CO_2 as substrate for growth (Ribbe et al., 1997).

Genetic organization of the nif complex of A. vinelandii

The major *nif* cluster from *A. vinelandii* encodes 15 *nif* specific genes whose products bear significant structural identity to the corresponding *nif* specific gene products from *K. pneumoniae* (Jacobson *et al*., 1989a). These genes include *nifH, nifD, nifK, nifT, nifY, nifE, nifN, nifX, nifU, nifS, nifV, nifW, nifZ, nifM,* and *nifF* (Fig. 1.3). Twelve other potential genes whose expression could be subject to *nif* specific regulation are interspersed among the identified *nif* specific genes (Fig. 1.3). These potential genes do

not encode products that are structurally related to the identified *nif*-specific gene products. Eleven potential *nif*-specific promoters were identified within the major *nif* cluster, and nine of these are preceded by an appropriate upstream activator sequence. Although there are significant spatial differences, the identified *A. vinelandii nif* specific genes have the same sequential arrangement corresponding to *nif* specific genes from *K. pneumoniae* (Jacobson *et al*., 1989a).



The black arrows indicate approximate position and direction of identified or proposed *nif* specific promoters. The genes whose deduced products have sequence identity when compared with *K. pneumoniae nif* specific gene products have the appropriate *nif* genotypic designations

Fig. 1.3. Organization of the major *nif* cluster from *A. vinelandii*.

The same study that determined the physical and genetic map of the major *nif* gene cluster from *A. vinelandii* also investigated the effect of various *nif* mutants on diazotrophic growth (Jacobson *et al*., 1989a). This and other structural-functional studies were useful in determining the characteristics of proteins encoded by the *A. vinelandii nif* region, as described below.

(i) NifH

The *nifH* gene codes for the dinitrogenase reductase or the Fe protein. The structure of the Fe protein and its role in nitrogen fixation has been described in details separately in this text.

(ii) NifD, NifK

The *nifD* and *nifK* genes code for the á and â subunits of the dinitrogenase or the MoFe protein, respectively. The structure of the MoFe protein and its role in nitrogen fixation has also been discussed separately.

(iii) NifT, NifY/NafY

The putative protein encoded by *nifT* gene has been deduced only from the DNA sequence. The function of this protein is completely unknown. The existence of conserved open reading frames and their linkage to the *nifHDK* genes indicates a possible role in nitrogenase function (Jacobson et al., 1989a). However, since the phenotypic characterization of *nifT* and *nifT*-overexpressing *K. pneumoniae* strains for effects on the regulation, maturation, and activity of nitrogenase identified no properties that were distinct from those of the wild type, it was concluded that *nifT* was not essential for nitrogen fixation under the conditions examined (Simon et al., 1996). The *nifY* is one of the few *nif* genes that has not been identified in many other nitrogen fixing organisms. The function of *nifY* gene product also remains unknown though it is suggested to aid in the insertion of FeMo cofactor into apodinitrogenase in *K. pneumoniae* (Homer et al., 1993). The *A. vinelandii* NafY protein (nitrogenase accessory factor Y, also known as □), although not located in the *nif* genetic region, was reported to be a molecular chaperone that assisted in maintaining the apodinitrogenase in a conformation that

facilitated the insertion of the FeMoco by acting as a metallo-insertase (Rubio et al., 2002). Mutational analysis of the strains containing mutations in both *nafY* and *nifX* showed that these were severely affected in diazotrophic growth, indicating their physiological importance in the stabilization of the apodinitrogenase (Rubio et al., 2002). Subsequently, the 3-dimensional structure of the NafY protein from *A. vinelandii* revealed two distinct domains of this protein: an N-terminal (Met1 to Leu98) domain and a C-terminal (Glu99 to Ser232) 'core' domain (Dyer et al., 2003). The NafY core domain was shown to be capable of binding the FeMoco of nitrogenase but unable to bind to apodinitrogenase in the absence of the N-terminal domain (Dyer et al., 2003).

(iv) NifE, NifN

Sequence similarity of *nifE* and *nifN* to *nifD* and *nifK* has led to the hypothesis that NifE and NifN form a complex that is structurally homologous to the MoFe protein (Brigle et al., 1987). It could therefore be speculated that the NifEN contains two metalloclusters sites, similar to those in the NifDK protein. One such site within the NifEN complex may be analogous to the MoFe protein P-cluster site, whereas the other site may provide a place for the FeMoco assembly (Goodwin et al., 1998). The proposed P-cluster analogue in NifEN has been identified as a [4Fe-4S] cluster that is likely coordinated at the NifE–NifN interface by NifE–Cys-37, NifE–Cys-62, NifE–Cys-124, and NifN–Cys-44 (Goodwin et al., 1998). A FeMoco precursor synthesized in the presence of *nifB* gene product was identified in studies by Hu et al. (Hu et al., 2005) and this NifEN-bound precursor contained Fe as the only metal, and Mo and homocitrate were added while the cluster was still bound to the NifEN complex or at a later step (Hu

et al., 2005). More recently, the NifEN-bound precursor was found to be a molybdenum-free analog of FeMoco and not one of the more commonly suggested cluster types based on the standard [4Fe-4S] architecture (Corbett et al., 2006). The NifEN fusion protein generated by Suh et al. was shown to be able to retain its functionality, thus indicating that these metalloproteins could accommodate gene organization and structural changes to a wide extent.

(v) NifX

The *nifX* gene encodes a protein of molecular mass 17 kDa. In *A. vinelandii* and *R. capsulatus,* the *nifX* product was shown to be required for *in vitro* FeMo cofactor synthesis (Shah et al., 1999). The exact role played by NifX in cofactor biosynthesis is not known. As the *nifX* gene is located in the same transcriptional unit as *nifE* and *nifN*, it has been proposed that it may play a role in the transfer of the cofactor precursor from the NifNE complex to the NifH or another protein, where further steps of FeMo cofactor biosynthesis are completed. Previous studies in *Azotobacter vinelandii* have identified certain important molecular interactions of the NifX protein with the other Nif proteins involved in the FeMoco biosynthesis/transport pathway; mainly it was shown that NifX binds to the NifB-co and also a FeMoco precursor from the NifEN complex (Shah et al., 1999; Rangaraj et al., 2001).

(vi) NifU, NifS

The NifU and NifS are involved in the acquisition of iron and sulphur necessary for the maturation of the Fe-S clusters found in both the Fe and the MoFe proteins. NifU is a dimer containing a 2Fe:2S clusters in each subunit and is required for the complete

activation or the catalytic stability of the nitrogenase Fe protein. NifS is a homodimeric cysteine desulfurase that supplies the inorganic sulfide necessary for formation of the Fe-S clusters contained within the nitrogenase component proteins (Zheng et al., 1993). It has been reported that NifS also catalyzes the removal of selenium from selenocysteine thus leading to the postulation of a mechanism similar to the L-cysteine reaction (Zheng et al., 1994). It has been suggested that NifU complements NifS either by mobilizing the Fe necessary for nitrogenase Fe-S cluster or by providing an intermediate Fe-S cluster assembly site (Dos Santos et al., 2004).

(vii) NifB, NifQ

The genes *nifB* and *nifQ* constitute a transcriptional unit in *K. pneumoniae* (Buikema et al., 1987). In *A. vinelandii* these genes are located outside of the *nif* gene cluster, constituting together with *nifLA*, a second unlinked *nif* gene cluster (Jacobson et al., 1989a). The *nifBQ* transcriptional unit of *A. vinelandii* has been shown to be required for the activity of the Nif, Anf and Vnf nitrogenase systems (Joerger et al., 1988). NifB-co, the presumed metabolic product of NifB, is an [Fe–S] cluster of unidentified structure that serves as a precursor to FeMo-co as well as to the FeV and FeFe cofactors (Shah et al., 1994). In *A. vinelandii* the NifQ(-)-producing mutation resulted in an increased (approximately 1,000-fold) Mo requirement for Mo nitrogenase activity (Rodriguez-Quinones et al., 1993).

(viii) NifV, NifW, NifZ

The *nifW* and the *nifZ* gene products are required for the full activation or catalytic stability of the MoFe protein (Jacobson et al., 1989b). Previous studies of *nifW*

and *nifZ* mutations in *A. vinelandii* showed that when either of these two genes were mutated, the wild type nitrogenase activity was affected to similar extents (Jacobson et al., 1989b). Genetic analysis of the NifW utilizing the yeast two-hybrid system revealed that the NifW of *Azotobacter vinelandii* interacts with the NifZ to form higher-order complexes (Lee et al., 1998). NifV is a homodimer that catalyzes the condensation of acetyl coenzyme A (acetyl-CoA) and alpha-ketoglutarate (Hoover et al., 1987). Also, the deletion of the *nifV* gene was shown to result in lower MoFe protein activity, suggesting the accumulation of an altered FeMo cofactor (Hawkes et al., 1984).

(ix) NifM

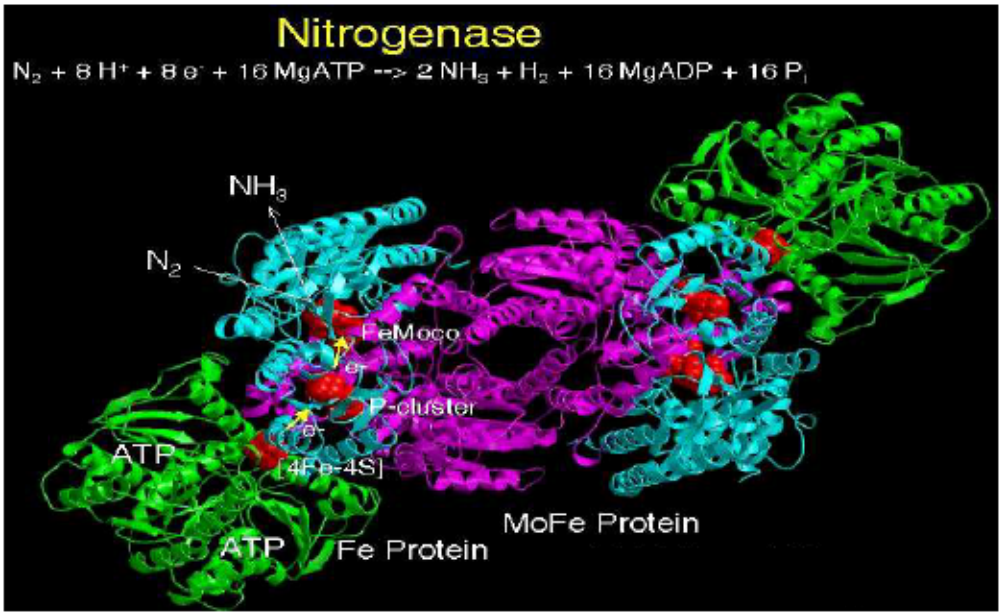
The *nifM* gene product is thought to activate the Fe protein via a modification or processing step that does not alter the mobility of the Fe protein subunits under denaturing conditions as shown by Roberts et al. in *K. pneumoniae* (Roberts et al., 1978). Based on homology analyses, the NifM has been proposed to be a PPiase of the parvulin family (Gavini and Pulakat, 2002). Very recently, genetic complementation studies of Human Pin1 in *A. vinelandii* revealed that the amino terminus of the NifM was required to deliver the PPiase effect to the NifH protein (Raja et al., 2006). Also, the nature of the molecular signature that defines the NifM dependence of NifH was investigated by screening a library of *nifH* mutants in *A. vinelandii* that acquired NifM independence (Gavini et al., 2006). They reported that NifH could acquire NifM independence by the replacement of the conserved Pro258 located in the C-terminal region of NifH with serine (Gavini et al., 2006).

Structural features of the nitrogenase enzyme

The well-studied Mo-nitrogenase consists of two component metalloproteins designated the iron protein (Fe protein) and the molybdenum iron protein (MoFe protein) that catalyze the ATP- dependent reduction of dinitrogen to ammonia (Fig. 1.4). The nitrogenase complex is extremely sensitive to inactivation by O_2, probably owing to the metallocenters that contain iron in them.

(i) The MoFe protein

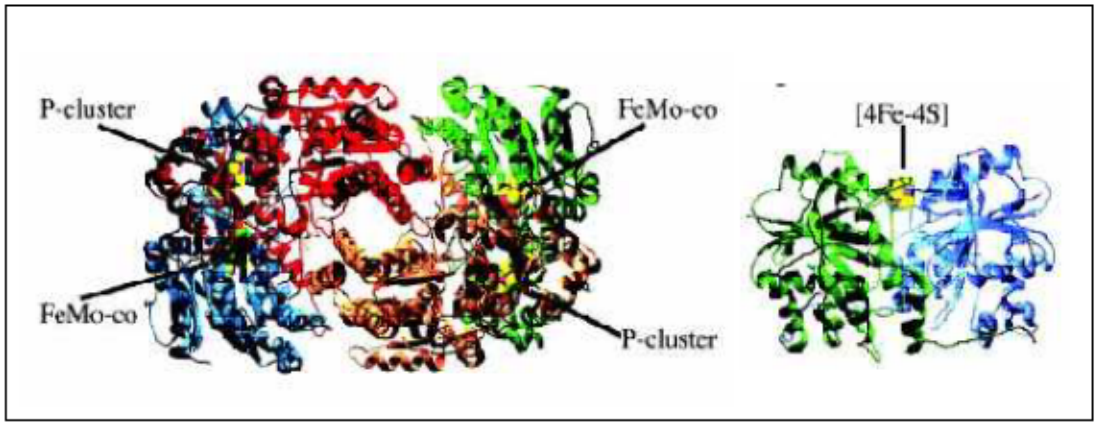
The MoFe-protein contains the active site for substrate reduction, and is organized as an □$_2$□$_2$ tetramer of molecular weight approximately 240kDa (Fig. 1.5). Associated with



The green ribbons represent the Fe protein component (NifH); the cyan and purple ribbons represent the NifD and NifK subunits of the MoFe protein component respectively. The [4Fe-4S] metallic center of the Fe protein, P-cluster and FeMoco centers of the MoFe protein are shown in red and the path of electron transfer during the nitrogen fixation reaction is indicated by yellow arrows (Schindelin et al., 1997).

Fig. 1.4. Structure of the nitrogenase enzyme complex.

this protein are 2 Mo, 30 Fe and 32 S organized into two copies each of two unique metalloclusters designated the FeMo-cofactor and the P-cluster. The FeMo-cofactor (or FeMoco) represents the site of substrate reduction, while the P-cluster is probably the initial acceptor of electrons from the Fe-protein.



The respective metallic clusters of both proteins are indicated (Rubio and Ludden, 2005).

Fig. 1.5. Structure of the MoFe protein (left) and Fe protein (right) components of the nitrogenase complex.

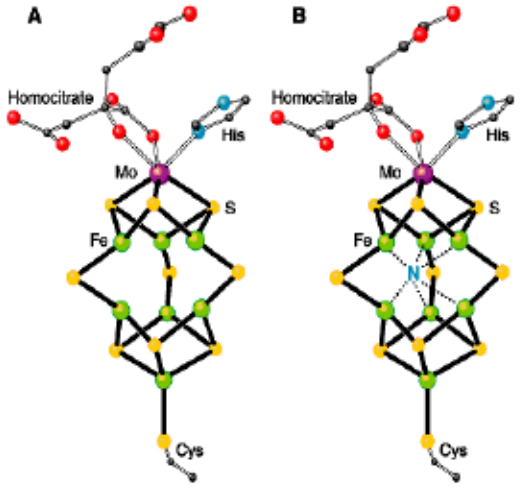
The Fe-protein mediates the coupling of ATP hydrolysis to electron transfer and is the only known electron donor that can support substrate reduction by the MoFe-protein. Although there is minimal amino acid sequence homology between the □ and □ subunits of the *A. vinelandii* MoFe protein, they are known to exhibit similar polypeptide folds, which consist of three domains of the parallel □-sheet / □-helical type (Kim and Rees, 1992a; Kim and Rees, 1992b). At the interface between the three domains is a wide, shallow cleft; in the □ subunit, the FeMo cofactor occupies the bottom of this cleft.

The P cluster is buried at the interface between a pair of □ and □ subunits with a pseudo-2-fold rotation axis passing between the two halves of the P cluster and relating the two subunits (Kim and Rees, 1992a). The extensive interaction between □ and □ subunits in an □□ dimer suggests that they form the fundamental functional unit. An open channel of ~8 Å diameter exists between the two pairs of □□ dimers with the tetramer 2-fold axis extending through the center (Kim and Rees, 1992a). The tetramer interface is dominated by interactions between helices from the two □ subunits, along with a cation binding site, presumably occupied by calcium, that is coordinated by residues from both □ subunits (Howard and Rees, 1996; Blanchard and Hales, 1996).

Structural models for the nitrogenase FeMo-cofactor and P-clusters

Each □□ unit of the MoFe protein contains an [8Fe-7S] cluster and a [7Fe-9S-Mo-homocitrate] cluster, respectively designated the P-cluster and FeMo-cofactor (Kim and Rees, 1992a). Structural models for the nitrogenase FeMoco and P-clusters were first proposed based on crystallographic analysis of the nitrogenase MoFe protein from *Azotobacter vinelandii* at 2.7 angstrom resolution (Kim and Rees, 1992a). These structures were also verified at 2.2 Å resolution (Chan et al., 1993). The 2.2 Å resolution analyses of the P-cluster pair confirmed the presence of eight Fe atoms in an arrangement consisting of two Fe_4S_4 cubane clusters linked by bridging Cys thiolate groups. The study also established that the two cubane clusters were further joined by a disulfide bridge between the S atoms in each Fe_4S_4 cubane cluster. The presence of the disulphide bond in the P-cluster pair suggested that it could act as a two-electron redox group,

involving cleavage and reformation of the disulfide bridge (Chan et al., 1993). The P-clusters are liganded by six cysteine thiol groups, two of which bridge the two clusters, □88 and □95, and four which singly coordinate the remaining Fe sites, □62, □154, □70, and □153 (Kim and Rees, 1992b). The FeMoco is a heterometallic double cubane consisting of one [4Fe-3S] and one [Mo-3Fe-3S] partial cubane bridged by three sulfides that share a μ_6-central atom, the identity of which is yet unknown but speculated to be C, O, or N.



Model of the FeMo cofactor of the nitrogen-fixing enzyme nitrogenase (A) before and (B) after the Einsle *et al.* report (Einsle et al., 2002). This report presented a high-resolution structure of the MoFe protein of nitrogenase, which contains the FeMo cofactor. A previously unrecognized interstitial atom in the FeMo cofactor that may possibly be nitrogen (blue N) was proposed in this new report. Carbon, gray; iron, green; sulfur, yellow; molybdenum, purple; oxygen, red; and nitrogen, blue.

Fig. 1.6. Nitrogen in the middle (Einsle et al., 2002).

Situated entirely in the □-subunit, the FeMoco is coordinated to the MoFe protein through the side chains of Cys275 and His442 which are bound to the Fe and Mo sites

located at opposite ends of the FeMoco (Kim and Rees, 1992a). The Mo site exhibits approximate octahedral coordination geometry and is bound to three sulfurs in the cofactor, two oxygens from homocitrate, and the imidazole side chain of His □442 (Kim and Rees, 1992b). Certain studies have examined a form of the MoFe protein that is missing the FeMo-cofactor (called apo-MoFe protein) (Christiansen et al., 1998; Ribbe et al., 2002). The apo-MoFe protein was isolated from either a □*nifB* or a □*nifH* strain of *A. vinelandii*. Deletion of either of these genes resulted in disruption of the biosynthesis of the FeMo-cofactor. A recent crystal structure of the MoFe protein revealed that the FeMoco contains a light atom as a central ligand (Einsle et al., 2002) (Fig. 1.6). Even at the high resolution of 1.16 Å, the central atom could not be confirmed, but the possibilities could be limited to nitrogen, oxygen, or carbon. The studies suggested that the central ligand was probably a nitrogen atom because of the presence of nitrogen as a substrate.

In a theoretical study it was again shown that the central ligand was most probably a nitrogen atom (Hinneman and Norskov, 2003) and other theoretical studies had similar conclusions (Lovell et al., 2003). In order to determine the pathway through which the FeMoco is inserted in the final step of the MoFe protein assembly, the structure of the FeMoco-deficient MoFe protein from an *Azotobacter vinelandii* □*nifB* strain was solved to 2.3 Å resolution; comparison with the wild type Fe protein showed that the residue rearrangements occurring in domain □III (□345 to □480) were involved in creating a funnel that allowed the entry of the FeMoco into the MoFe protein (Schmid et al., 2002). Within the □III domain, the loop from residues □353 to □364 was located at the entrance

of the funnel. Additionally Lys (□315 and □426), Arg (□96, □97, and □277), and His (□274, □442, and □451) residues in the funnel area were involved in creating a positively charged environment for the FeMoco insertion (Schmid et al., 2002).

(ii) The Fe protein

The Fe protein is a 64-kDa homodimer and contains a single [4Fe-4S] cluster coordinated between the two subunits (Georgiadis et al., 1992). The thiol ligand of C97 and C132 from each Fe protein subunit coordinates the cubane [4Fe-4S] cluster (Schlessman et al., 1998) (Fig. 1.7).

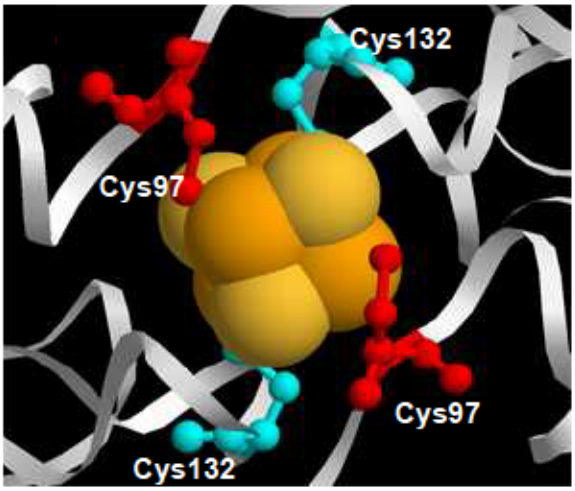
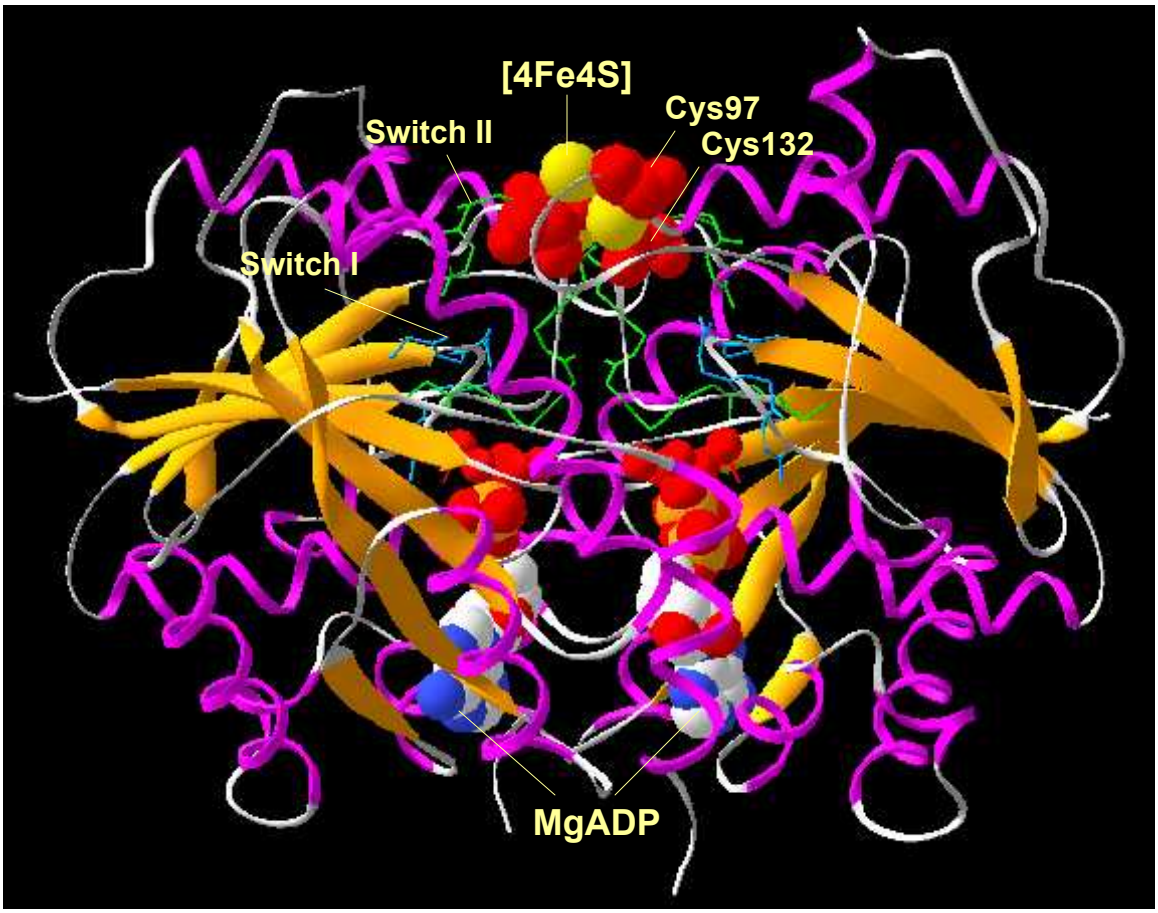


Image created by using the X-ray crystallographic structure of the Fe protein (PDB ID: 1FP6 (Jang et al., 2000)) from the protein database (www.rcsb.org)

Fig. 1.7. The [4Fe-4S] cluster of the Fe protein liganded centrally between two NifH monomers by Cys97 and Cys132 residues.

The Fe protein serves at least three different functions in the nitrogenase enzyme system (Allen et al., 1994) (i) It serves as an obligate electron donor to the molybdenum-iron (MoFe) protein component of the nitrogenase enzyme (ii) participates in the

biosynthesis of FeMo cofactor of the MoFe protein and (iii) helps in maturation of the apodinitrogenase to a FeMoco-active form.



The [4Fe-4S] metallic center and its Cys ligands, Walker motif A / MgADP binding site (residues 9-16), Switch I (residues 39-67 in blue) and Switch II (residues 125-132 in green) of the Fe protein are highlighted. Image generated from X-ray crystallographic structure of the Fe protein (PDB ID: 1FP6).

Fig. 1.8. Structural features of the Fe protein.

Based on mutagenesis and crystallographic studies, the Fe protein appears to consist of the following regions of significance: the Walker motif or P-loop, a nucleotide binding motif containing the G-X-X-X-X-G-K-S/T consensus sequence (residues 9 to 16

in *A. vinelandii* NifH), two switch regions, Switch I and Switch II, which interact with the Mg^{2+} of the bound nucleoside triphosphate, the [4Fe-4S] metal cluster ligands, and the domains involved in MoFe docking (Fig. 1.8). The Fe protein region corresponding to Switch I includes residues ~39 to 67 and the residues in Switch II are 125 to 132. The amino acids 125 to 128 of the *A. vinelandii* Fe protein form the conserved D-X-X-G sequence motif of Switch II. The switch regions appear to be involved in transduction of the information concerning the nucleotide binding status of the Fe protein and orchestrate conformational changes of the Fe protein necessary for nitrogen fixation (Jang et al., 2000).

Structural models for the [4Fe-4S] metal center of the Fe protein

Protein-bound [FeS] clusters function widely in biological electron-transfer reactions, and their midpoint potentials control both, the kinetics and thermodynamics of these reactions (Stephens et al., 1996). The Fe protein component of nitrogenase contains a [4Fe-4S] metallic center and two nucleotide binding sites, one in each subunit (Georgiadis et al., 1992). It is known that nucleotide binding causes changes in the midpoint potential of the [4Fe-4S] of the Fe protein, presumably by triggering changes in the protein environment around the cluster (Spee et al., 1998). These conformational changes result in a Fe protein-MoFe protein complex formation leading to electron transfer from the [4Fe-4S] cluster of the Fe protein to the substrate reduction site of the MoFe protein. The [4Fe-4S] cluster was shown to function as a hinge region between the two nucleotide binding domains participating in the cooperative binding of two

nucleotides (Ryle et al., 1996b). In this study, when the Ala98 residue of the *A. vinelandii* Fe protein (located near the [4Fe-4S] cluster) was changed to Val and Gly by site directed mutagenesis, it was observed that the Ala98Val Fe protein lost its cooperativity in binding a nucleotide (Ryle et al., 1996b).

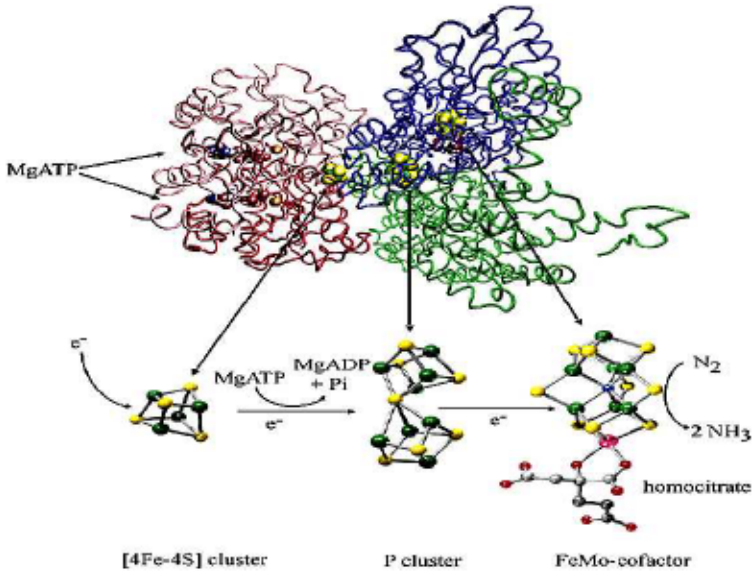
Mechanistic and functional aspects of the nitrogenase enzyme

The Fe protein and the MoFe protein work together as a molecular machine to catalyze the reduction of N_2. The Fe protein acts as a reductant of the MoFe protein, transferring one electron at a time from its [4Fe–4S] cluster to the MoFe protein in a reaction linked to the hydrolysis of MgATP, according to the equation:

$$N_2 + 8H^+ + 8e^- + 16\ ATP \rightarrow 2\,NH_3 + H_2 + 16ADP + 16\ Pi$$

The recent model for the catalytic mechanism of nitrogenase suggests that the reduced Fe protein, with two bound MgATP molecules, binds to the MoFe protein and a single electron is transferred from the Fe protein to the MoFe protein with the concomitant hydrolysis of two MgATP molecules (Fig. 1.9) (Thorneley & Lowe, 1983; Howard & Rees, 1994). The Fe protein then dissociates from its partner MoFe protein following each electron transfer event (Hageman and Burris et al., 1978) allowing the Fe protein to be recharged by reduction and replacement of the spent nucleotides with MgATP. A minimum of eight such association/dissociation events is required for each N_2 reduced. This oxidized Fe protein, with two molecules of MgADP bound, is then released from the MoFe protein, and another reduced Fe protein, with two bound MgATP molecules, binds to the partially reduced MoFe protein for a second round of MgATP hydrolysis and electron transfer (Hageman and Burris, 1978). Following this step, the

dissociation of the Fe protein-MoFe protein complex occurs. This dissociation step takes place after the transfer of each electron and is the rate-limiting step in nitrogenase catalysis. This cycle is repeated until sufficient electrons have been transferred to the MoFe protein to reduce the bound substrate



The Fe protein is shown on the left (identical subunits in pink and red), and one catalytic □□ dimer of the MoFe protein is shown on the right (□-subunit in blue and □-subunit in green). The associated metal clusters and MgATP located within the nitrogenase complex are shown as space-filling models. Electrons flow in an ATP-dependent reaction from the [4Fe-4S] cluster of Fe protein to the P cluster and FeMo-cofactor of the MoFe protein, where the reduction of N_2 to ammonia occurs.

Fig. 1.9. The direction of electron flow in the nitrogenase complex during nitrogen reduction (Dos Santos et al., 2004).

MgATP serves two known functions in the nitrogenase mechanism. Firstly, MgATP binding to the Fe protein triggers protein conformational changes which are reflected as changes in the properties of the [4Fe-4S] cluster (Mortenson et al., 1993). Consequently a significant lowering of the redox potential of the [4Fe-4S] cluster from -

296 to -420 mV takes place (Zumft et al., 1974; Watt et al., 1986). This lowering of the redox potential along with other changes induced in the protein could be necessary for the docking of the Fe protein onto the MoFe protein and subsequent electron transfer. Secondly, MgATP couples its hydrolysis to electron transfer from the Fe protein to the MoFe protein and to substrate reduction (Mortenson et al., 1993). A study by Lanzilotta et al. provided major evidence for (i) the tight binding of the Fe protein to the MoFe protein and (ii) electron transfer from the Fe protein to the MoFe protein, subsequent to MgATP induced conformational changes of the Fe protein (Lanzilotta et al., 1996). The study utilized a Leu127 deletion (L127□) created in the MgATP signal transduction domain of the Fe protein (Howard and Rees, 1994) that showed protein conformational changes characteristic of the MgATP-bound state even in the absence of MgATP. The L127□ Fe protein showed high affinity binding to the MoFe protein and furthermore, electron transfer from the altered Fe protein to the Fe protein was also observed, indicating events taking place in the nitrogenase complex in an MgATP-bound state (Lanzilotta et al., 1996). Very recently the structure of the nitrogenase L127□ Fe protein with MgATP bound was determined (Sen et al., 2006). In that, it was suggested that MgATP may enhance the stability of an open conformation and prohibit intersubunit interactions, which have been implicated in promoting nucleotide hydrolysis. This could be critical to the tight control of MgATP hydrolysis observed within the nitrogenase complex and may be important for maintaining unidirectional electron flow toward substrate reduction (Sen et al., 2006).

(A) *The Fe protein cycle*

The nitrogenase Fe protein is the only known reductant of the MoFe protein that helps in substrate reduction. As mentioned earlier, the reduction of the MoFe protein by the Fe protein occurs during the transient association between these two proteins, coupled with the hydrolysis of MgATP on the Fe protein. Based on the various properties of the Fe protein and the steps that involve the Fe protein in the process of nitrogen fixation, Igarashi et al. outlined the following 'Fe protein catalytic cycle' (Igarashi et al., 2003):

(i) Redox properties of the Fe protein

The oxidized state of the Fe protein, with the [4Fe–4S] cluster in the 2^+ oxidation state ($2Fe^{2+}$ and $2Fe^{3+}$), can be achieved through oxidation by various agents, including O_2, redox active dye mediators and electron transfer to the MoFe protein (Thorneley and Ashby, 1989; Lindahl et al., 1985). The [4Fe–4S] cluster of the Fe protein can also be found in the reduced all ferrous state with an overall oxidation state of (0) ($4Fe^{2+}$) (Angove et al., 1997; Watt and Reddy, 1994).

(ii) Nucleotide binding to the Fe protein

The Fe protein binds two equivalents of MgATP, one in each subunit (Mortenson et al., 1993). Similar to other nucleotide-binding enzymes, divalent metal ions are required for ATP binding. Mg^{2+} is the likely physiologically relevant metal for Fe protein, but other metals (*e.g.*, Mn^{2+}, Ca^{2+}, or Fe^{2+}) have also been shown to enable nucleotide binding (Weston et al., 1993). During initial research on nitrogenase, it was found that the binding of nucleotides to the Fe protein caused changes in the properties of the protein, including the properties of the [4Fe–4S] cluster (Bui and Mortenson, 1968). It was speculated that specific segments of the Fe protein were involved in the

communication of a signal from the MgATP binding site to the [4Fe-4S] so as to effect conformational changes in the Fe protein. This region was most likely predicted to be the peptide section of the Fe protein leading from the MgATP binding site (125^{Asp}) to the [4Fe–4S] cluster ligand (132^{Cys}) (Howard and Rees, 1994).

(iii) Fe Protein-MoFe protein complex formation

Several important events in the Fe protein cycle are initiated upon the docking of the Fe protein on the MoFe protein. Through chemical cross-linking studies, Willing et al. revealed the docking surface on the Fe protein to be near the [4Fe–4S] cluster and on the MoFe protein near the □-□ interface (Willing and Howard, 1990). The biochemical and structural characterization of the cross-linked complex by Schmid et al. showed that the cross-linking reaction was inhibited by glycinamide which was incorporated in the Glu112 position of the Fe protein (Schmid et al., 2002). Crystallographic analysis of the EDC-cross-linked complex at 3.2 Å resolution confirmed the site of the isopeptide linkage (Schmid et al., 2002). One of the possible effects of peptide conformational changes around the [4Fe-4S] cluster is a decrease in its redox potential (Ryle et al., 1996a, Kurnikov et al., 2001). Also, upon formation of the Fe protein-MoFe protein complex, the [4Fe-4S] cluster and P-clusters become better reductants and favor electron transfer to the FeMoco active site (Kurnikov et al., 2001).

(iv) MgATP hydrolysis and electron transfer

Although MgATP hydrolysis and electron transfer influence each other, they are not obligatorily dependent. Both MgATP hydrolysis in the absence of electron transfer and electron transfer in the absence of MgATP hydrolysis have been demonstrated

(Cordewener et al., 1988). Subsequent to the Fe protein docking onto the MoFe protein, MgATP hydrolysis in the Fe protein is activated and electron transfer from the Fe protein to the MoFe protein occurs. Based on the observation that the substitution of either of the two surface Phe residues on the MoFe protein (□–125^{Phe} or □–125^{Phe}) by Ala caused a decrease in proton reduction activity and substitution of both by Ala stopped electron transfer, the Phe residues were thought to be involved in activating the trigger on the Fe protein and initiating MgATP hydrolysis (Christiansen et al., 2000). The increase in rate of electron transfer effected by MgATP hydrolysis may be as a result of (a) increasing the difference in redox potentials of the redox pair (□E_m), (b) decreasing the distance between the [4Fe–4S] cluster of the Fe protein and P-cluster of the MoFe protein, or (c) altering the intervening pathway to favor electron transfer, or some combination of the above (Igarashi et al., 2003). The rate of electron transfer is certainly favored by a shift of the [4Fe–4S] cluster in the Fe protein ~5 Å closer to the P-cluster, triggered by docking of the Fe protein on the MoFe protein (Schindelin et al., 1997).

(v) Dissociation of the Fe protein from the MoFe protein

In the nitrogenase reaction, the dissociation of the Fe protein from the MoFe protein following MgATP hydrolysis and electron transfer is probably the rate limiting step (Wilson et al., 2001). The need for dissociation of the component proteins as a prerequisite to each electron transfer event is supported by the observation that the Leu127 deletion variant (□127Leu) Fe protein, when complexed to the MoFe protein, can transfer a single electron to the MoFe protein, but cannot perform subsequent electron transfers (Lanzilotta et al., 1996).

(B) *Events in the MoFe protein*

As noted earlier, the MoFe protein component of nitrogenase contains the active site metal center known as FeMoco, and the P-cluster. The $\alpha_2\beta_2$ MoFe protein can be thought of as two catalytic $\alpha\beta$-units, and each $\alpha\beta$-unit contains one P-cluster and one FeMo-cofactor. The general mechanism of nitrogenase includes Fe protein docking on the MoFe protein near the P-cluster, transfer of an electron to the P-cluster and the transfer of one or more electrons from the P-cluster to FeMo-cofactor, where substrates are reduced.

(i) Catalytic role of the P-cluster

Some studies have pointed to a redox role of the P-cluster during nitrogenase catalysis (Ma et al., 1996; Gavini et al., 1994). When the MoFe protein is isolated with excess dithionite, the P-cluster is in the P^N state where all of the Fe atoms are essentially ferrous (Zimmermann et al., 1978). From the P^N state, the P-cluster can be oxidized to many more states using oxidative dyes (Tittsworth and Hales, 1993). According to the consensus model, the P-cluster accepts electrons from the Fe protein and donates electrons to FeMo-cofactor. However, the step involving the P-cluster probably occurs transiently, since the FeMoco is reduced very rapidly subsequent to the electron transfer from the Fe protein to the MoFe protein and also because the intermediate state has not been clearly detected (Smith et al., 1973). The topic of the role of the P-cluster in electron transfer, and the pathway of transfer to FeMoco still remains largely under investigation.

(ii) Mechanism of substrate binding and reduction

The reduction of dinitrogen and other substrates by the MoFe protein requires two or more rounds of the Fe protein cycle. Reactions such as proton or acetylene reduction require an accumulation of two electron equivalents in the MoFe protein, whereas reduction of dinitrogen requires 8 electron equivalents. Although it is obvious that the MoFe protein accumulates electrons for these different reactions, it is not known where and how these electrons are accumulated. Some possible sites of accumulation include the P-cluster, FeMo-cofactor, bound hydrides, and partially reduced bound substrate intermediates (*reviewed in* Igarashi et al., 2003). Some theoretical studies indicate that nitrogen could have many different binding modes, such as perpendicular or parallel to the length of the FeMoco or binding with two to four Fe atoms or binding end-on or side-on or within the central cavity of the FeMoco (Dance, 1996). A recent study focused on the preparatory migration of hydrogen atoms around the FeMo-co active site of nitrogenase and suggested that binding sites and binding states for substrates and intermediates are characterized not only by their locations on the FeMoco face but also by the structural and temporal status of the distribution of H atoms over the FeMo-co reaction domain (Dance, 2006).

Maturation of the nitrogenase enzyme

Investigations into the synthesis of the FeMoco have shown that it is separately synthesized first and then inserted into the apo-MoFe protein (Ugalde et al., 1984, Robinson et al., 1986). However even active research has not provided all answers regarding the identity of proteins involved in this process, the exact sequence of

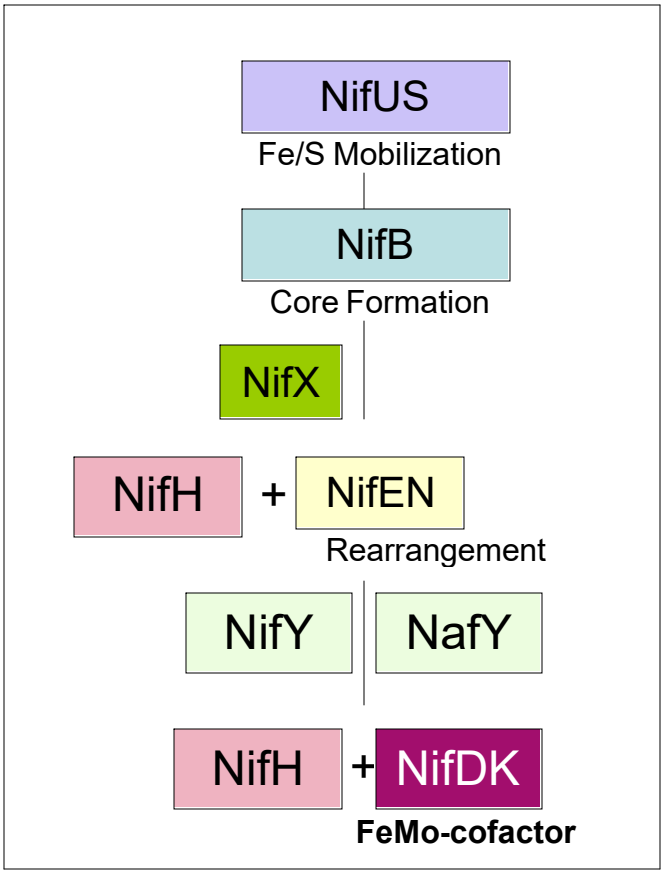
events or the specific roles of the individual players. The known biosynthetic pathway proteins and their roles are summarized in Table 1.2. Four important features have been observed with respect to the FeMo-cofactor biosynthesis (Dos Santos et al., 2004): (1) formation of an Fe-S core, (2) rearrangement of the Fe-S core to form an entity that is similar to the metal-sulfur core of FeMo-cofactor, (3) insertion of Mo and attachment of homocitrate, and (4) transfer of FeMo-cofactor or its precursors among the various sites at which these events occur. The flow of the Fe and S through the biosynthetic pathway has been proposed to be from NifUS →NifB →NifX →NifEN →NifY/Gamma →MoFe protein (NifDK) (Fig. 1.10).

Regulation of nitrogen fixation

Although similar regulatory networks are used to control nitrogen fixation, there are considerable differences in the regulatory networks from species to species (Dixon and Kahn, 2004). Most *nif* genes in the Proteobacteria are activated by the enhancer-binding protein NifA together with the RNA polymerase sigma factor σ^{54}. The expression of NifA and, in many cases its activity, is controlled by regulatory cascades that are responsive to different environmental signals (Dixon and Kahn, 2004). The *nifLA* operon codes for the regulatory proteins which control expression of all other nitrogen fixation operons. NifA is a transcriptional activator which is synthesized in an active form. Its activity is inhibited by NifL in response to nitrogen availability and oxygen tension (Hill et al., 1981, Merrick et al., 1982). The *ntr* system modulates nitrogen metabolism in

Table 1.2. *nif/naf* gene products involved in the overall FeMo-cofactor biosynthesis (Dos Santos et al., 2004)

Gene	product/function(s)
nifH	Fe protein subunit
nifD	MoFe protein □-subunit
nifK	MoFe protein □-subunit
nifB	involved in the production of an Fe/S-containing FeMo cofactor precursor, designated NifB-cofactor
nifQ	involved in FeMo-cofactor biosynthesis, probably at an early step
nifV	homocitrate synthase
nifX	probably an intermediate carrier in FeMo-cofactor biosynthesis
nifY	probably an intermediate carrier in FeMo-cofactor biosynthesis
nifN	subunit of $NifN_2E_2$; appears to provide a transient site upon which one or more events related to FeMo-cofactor assembly occur
nifE	subunit of $NifN_2E_2$
nifU	complements NifS in the mobilization of Fe and S for metallocluster assembly; required for the synthesis of active Fe protein and MoFe protein
nifS	pyridoxal-dependent cysteine desulfurase; required for the synthesis of active Fe protein and MoFe protein
nifW	required for the synthesis of a fully active MoFe protein
nifZ	required for the synthesis of a fully active MoFe protein
nifT	function unknown
nifF	flavodoxin
nifJ	pyruvate:flavodoxin oxidoreductase
nifA	positive regulatory element
nifL	negative regulatory element
nafY	probably an intermediate carrier in FeMo-cofactor biosynthesis



Flowchart diagram of the proposed pathway for the FeMo-cofactor. The boxes represent the proteins involved in this process, and the arrows represent the transfer of precursor forms through the process (Dos Santos et al., 2004).

Fig. 1.10. Proposed model for FeMo-cofactor biosynthesis

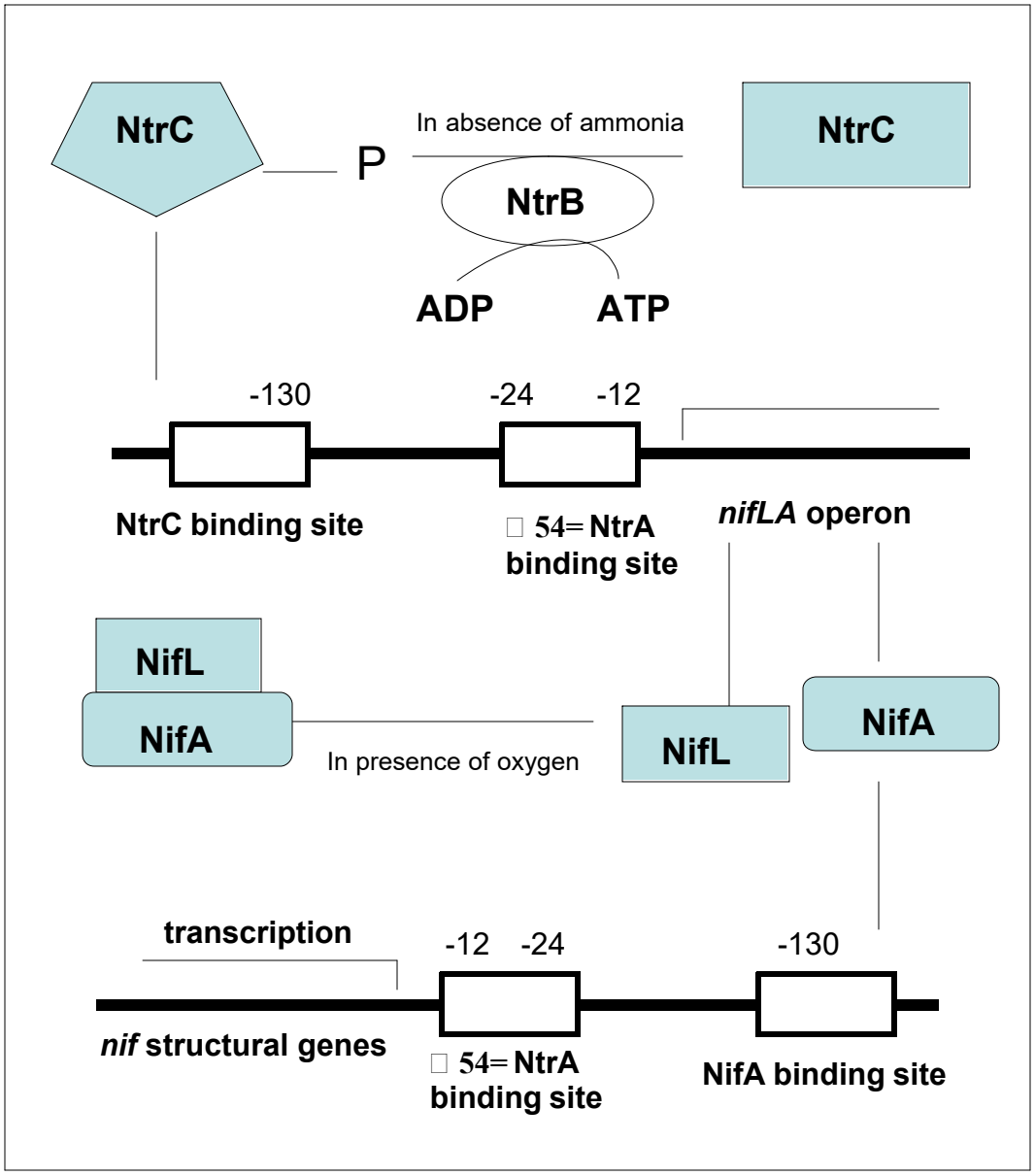
response to the prevailing nitrogen source and the requirements of the cell (Fig. 1.11). This system is composed of four enzymes: a uridylyltransferase/uridylyl removing enzyme (UTase/UR) encoded by the *glnD* gene, a small trimeric protein PII, encoded by the *nif* genes except the regulatory genes *nifLA* themselves. If the NifA protein is synthesized, its function is to activate the other *nif* genes. When oxygen is present, the NifL protein binds to NifA and prevents it from activating the other *nif* genes.

glnB, and a two component regulatory system composed of the histidine protein kinaseof the *nifLA* operon and activates transcription. NtrA (= GlnF = RpoN = □54) is the nitrogen sigma factor, which is needed for expression of the *nifLA* operon and the *nif* structural genes. NtrA is an alternative sigma factor used by RNA polymerase to recognize many genes involved in nitrogen metabolism which are not recognized by the standard sigma factor. The *nifA* gene encodes a protein required for switching on all of NtrB and the response regulator NtrC (Merrick and Edward, 1995). In many diazotrophs, an increase in nitrogen status has been shown to inactivate nitrogenase by the reversible ADP-ribosylation of a specific arginine residue in the Fe protein. In this process, the dinitrogenase ADP-ribosyl transferase (DraT) adds an ADP-ribose moiety and thereby inactivates the enzyme complex and dinitrogenase reductase activating glycohydrolase (DraG) removes the moiety and thus activates the enzyme (Ludden et al., 1988, Masepohl et al., 1993).

Current challenges in the study of biological nitrogen fixation

The challenges that now face BNF research include the enhancement of BNF in legumes and the transfer of important BNF traits to non-nitrogen fixing organisms. A major emphasis is also on making the BNF process a popular and agronomically efficient practice. Approaches must also be planned for the mobilization, redistribution and utilization of stored N reserves within host plants (Shantharam et al., 1997). It would also be very useful to obtain designer plants that could better utilize the plant's capacity and ability to improve nitrogen economy, maintain a favorable C/N ratio and tolerate

various ecological or environmental problems (Shantharam et al., 1997). It is also necessary to study the phylogeny and distribution of the wide range of nitrogen fixing organisms to understand newer evolutionary aspects of nitrogen fixation and nitrogenase systems. At the molecular level, the structural and accessory components of the nitrogenase system require more research in terms of understanding their structural design, working and function (Board on Science and Technology for International Development, 1994). Most importantly, the biosynthesis and functional aspects of the FeMoco need extensive analysis since this metallic cluster forms the catalytic site for nitrogen reduction. Next, it is essential that for creating transgenic plants that possess the ability to fix nitrogen, the *nif* system consisting of a large genetic region is condensed such that only the minimal genes required for conferring the ability to fix nitrogen have to be transferred into the plant. The process of nitrogen fixation is very energy demanding and it is thus highly regulated. Efforts could therefore be directed towards overcoming the inhibition of the nitrogenase system that occurs from even low levels of ammonium ions, thus helping to obtain higher yields of fixed nitrogen. It is also important to understand the reasons for the oxygen sensitivity of the nitrogenase enzyme and develop an oxygen insensitive nitrogenase so that it could better facilitate the transfer of nitrogen fixing ability to plants. Further, the challenge of decreasing the hydrogen



The *nif* genes are regulated by the *nifLA* operon. The nitrogen regulators NtrC (= GlnG), and NtrB determine whether or not the *nifLA* operon is expressed (depending on the presence of ammonia or organic nitrogen). In the absence of ammonia or organic nitrogen the NtrC protein is phosphorylated by the NtrB protein. NtrC-P then binds to the upstream region

Fig. 1.11. Nitrogen metabolism regulation by the *ntr* system.

formation during the nitrogen fixation reaction by genetic manipulation should be addressed, since such a step could lead to 25% more energy availability for the production of fixed nitrogen alone. Finally, newer technological and bioinformatics tools should be utilized to deal with questions concerning the metabolic and functional pathways of nitrogen fixation and thus improve our current knowledge of this intriguing system.

Aims and objectives of this study

Azotobacter vinelandii is a free-living, aerobic diazotrophic bacteria. It possesses the nitrogenase enzyme system that catalyzes the reductive conversion of nitrogen to ammonia. Several studies have utilized this organism for research related to biological nitrogen fixation owing to its ease of maintenance and growth, high recombination frequency at the genetic level and its physiological properties that allow it to grow aerobically and yet maintain an oxygen sensitive nitrogenase enzyme. In this study, we mainly dealt with the following research aims pertaining to the Mo-nitrogenase system of *A. vinelandii* : (i) Detection of protein-protein interaction between fused NifD-K proteins, mainly to test whether the NifD-K fusion protein functioned as a single unit or a dimer (ii) Investigation of the structural analogy between the NifDK and ChlBN proteins of the nitrogenase and light-independent (dark operative) protochlorophyllide reductase (DPOR) systems respectively (iii) Defining the role of the NifX protein in the FeMoco biosynthesis/insertion pathway using protein-protein interaction studies and (iv) Structural analyses of the NifH protein and its ability to complement the function of the

structurally analogous ArsA protein involved in arsenite resistance. All of these objectives are described in further details below.

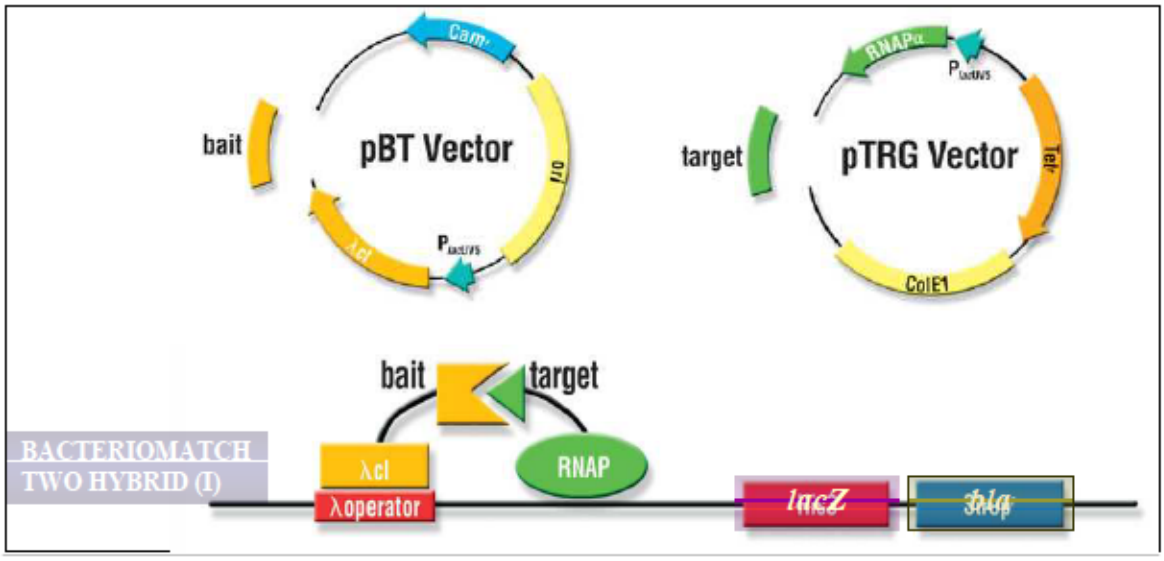
(i) A functional NifD-K fusion protein was obtained by Suh et al. previously in our lab (Suh et al., 2003). This was a major step towards the reduction of genetic material to facilitate the cloning of *nif* genes into plants. We were interested in determining the protein-protein interaction between the fusion NifD-K protein units to understand the functioning of the fusion protein. Therefore we utilized the BacterioMatch™ Two Hybrid system to examine if any interaction existed between two NifD-K fusion proteins and if present, whether the extent of interaction was similar to the native interaction between the NifK subunits.

BacterioMatch Two-Hybrid System

The BacterioMatch two-hybrid system is a genetic technique for detecting protein-protein interaction. The two fusion proteins -- bait and target are co-expressed in the reporter *E. coli* (XL1-Blue) cells and if these two proteins interact then they activate the expression of the reporter genes (Hu et al., 2000). The reporter genes allow for easy detection of the protein interactions either qualitatively or quantitatively.

The BacterioMatch two-hybrid system consists of two vector plasmids: the bait (pBT) plasmid and the target (pTRG) plasmid. The pBT plasmid is 3.2 kb in size and encodes the full length bacteriophage □CI protein under the control of the strong promoter lacUV5. At the 3' end of the □CI is a multiple cloning site (MCS) where any protein can be fused in the restriction sites. The bait plasmid carries a p15A replication origin and also a chloramphenicol resistance gene (BacterioMatch Instruction Manual,

Stratagene, La Jolla, CA). The pTRG plasmid is 4.4 kb in size and carries a ColE1 replication origin and tetracycline resistance gene. The plasmid directs the transcription of the amino terminal domain of RNA polymerase □□subunit is the multiple cloning site where the gene of interest can be fused (BacterioMatch Manual, Stratagene, La Jolla, CA). The XL1-Blue *E. coli* strain is used as the reporter strain in this technique. This strain lacks all the restriction systems and harbors *lacI*^{q} on the F' episome to repress the synthesis of bait and target proteins. The F' episome includes a reporter gene cassette containing *lacZ* (encodes □-galactosidase) and *bla* (encodes □-lactamase) reporter genes under the control of a modified *lac* promoter; this promoter has a single □ operator (OR2) centered at position -62 replacing the catabolite repressor protein (CRP) site originally associated with the *lac* promoter (BacterioMatch Manual, Stratagene, La Jolla, CA).



The *lacZ* and *bla* reporter gene expression levels indicate the extent of bait and target protein-protein interaction (BacterioMatch Manual, Stratagene, La Jolla, CA).

Fig. 1.12. Schematic representation of BacterioMatch Two-Hybrid System

As shown in Fig. 1.12, this system is based on transcriptional activation. A protein of interest in fused to the full length □CI protein in pBT and the corresponding target protein is fused to the N-terminal domain of the □-subunit of the RNA polymerase in pTRG. This is followed by cotransformation of the XL1-Blue reporter cells with both these plasmids. The N-terminal domain of the □CI protein in pBT is the DNA binding domain and has an affinity for the □ operator sequence and in this way the bait protein fused to □CI is also tethered to □ operator sequence of the reporter promoter on the F' episome. If there is an interaction between the bait and target proteins, the RNA polymerase is recruited and stabilized close to the promoter. In this way protein-protein interaction leads to the activation of transcription of the reporter genes *lacZ* and *bla*. The extent of gene expression can be measured by recording the □-galactosidase activity or ampicillin resistance.

(ii) We performed structural analyses of the ChlBN proteins, involved in protochlorophyllide reduction during photosynthesis, in comparison to the structurally homologous NifDK and NifEN protein of the nitrogenase system. This allowed us to

(iii) We used the BacterioMatch™ Two Hybrid system to detect protein-protein interactions between the NifX protein that has been indicated to be involved in the FeMoco biosynthetic pathway and other Nif proteins that participate in FeMoco biosynthesis/ insertion, such as NifB, NifH, NifN, NifD and NifK. Our studies were able to indicate the steps in the FeMoco biosynthetic pathway where the NifX protein may be involved.

(iv) The NifH protein structure and mechanism was compared to structurally related proteins such as MinD, ArsA, ChlL and CompA using structural modeling and bioinformatics tools. The ArsA protein is an ATPase similar to NifH but is involved in providing arsenite resistance. We were therefore interested in determining if the NifH protein had sufficient conservation of the structural domains necessary for performing the function of arsenite resistance when expressed as a chimeric NifH-ArsA2 protein, in which the N-terminal ArsA1 polypeptide of ArsA was replaced by the NifH. The phylogenetic relationship between ArsA and NifH was also derived to understand their divergence during evolution.

Apart from these, we also carried out preliminary investigations on the *A. vinelandii* transcriptome in comparison to the closely related *Pseudomonas aeruginosa* bacterium (Rediers et al., 2004) by hybridization of the *A. vinelandii* cDNA on a *P. aeruginosa* gene chip (Affymetrix® (Santa Clara, CA)). Currently, only the draft sequence of the *A. vinelandii* genome is available (Department of Energy, USA), and it is not clear by when the final annotated sequence would be published. Large scale gene expression studies of *A. vinelandii* may answer several perplexing questions regarding this bacterium but such studies have not been undertaken due to the lack of the complete genomic sequence. As an alternative, we therefore explored if it would be of purpose to utilize the *P. aeruginosa* Affymetrix® DNA chips to study the *A. vinelandii* transcriptome. We attempted to investigate, through such a gene expression study, the maintenance and protection of the oxygen sensitive nitrogenase in this aerobically respiring microorganism. Earlier reports have mostly indicated the high respiration rate

of *A. vinelandii* as the reason for scavenging oxygen away from the nitrogenase system (Poole and Hill, 1997). For this, we assessed the energy metabolism related genes in *A. vinelandii* in comparison to those of *P. aeruginosa* by analysis of their transcriptomic profiles and determined the significant differences. (The results of this gene expression study have been briefly described in the Appendix section of this dissertation since it was undertaken as a preliminary investigation only).

References

Allen, R. M., Chatterjee, R., Madden, M., Ludden, P. W., and Shah, V. K. 1994. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Crit. Rev. Biotech. 14: 225-249

Angove, H.C., Yoo, S.J., Burgess, B.K., and Münck, E. 1997. Mössbauer and EPR evidence for an all-ferrous Fe_4S_4 cluster with $S = 4$ in the iron protein of nitrogenase. J. Am. Chem. Soc. 119: 8730–8731.

Barney, B. M., Lee, H-I., Dos Santos, P. C., Hoffman, B. M., Dean, D. R. and Seefeldt, L. C. 2006. Breaking the N2 triple bond: insights into the nitrogenase mechanism. Dalton Trans. (19): 2277-2284

Blanchard, C. Z. and Hales, B. J. 1996. Isolation of two forms of the nitrogenase VFe protein from *Azotobacter vinelandii*. Biochemistry, 35(2): 472-478.

Board on Science and Technology for International Development, National Research Council (NRC). 1994. Biological nitrogen fixation: A review of research grants funded by the U. S. agency for international development, pp. 5-32. The National Academies Press, Washington, DC, USA.

Brigle, K. E., Weiss, M. C., Newton, W. E. and Dean, D. R. 1987. Products of the iron-molybdenum cofactor-specific biosynthesis genes, *nifE* and *nifN* are structurally homologous to the products of the nitrogenase molybdenum-iron protein genes, *nifD* and *nifK*. J. Bacteriol. 169(4): 1547-53

Bui, P. T. and Mortenson, L. E. 1968. Mechanism of the enzymatic reduction of N_2: the binding of adenosine 5'-triphosphate and cyanide to the N_2 reducing system. Proc Natl Acad Sci USA 61: 1021–1027.

Buikema, W. J., Klingensmith, J. A., Gibbons, S. L. and Ausubel, F. M. 1987. Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae nifB* genes. J. Bacteriol. 169(3): 1120-1126

Burgess, B. K. and Lowe, D. J. 1996. Mechanism of molybdenum nitrogenase. Chem. Rev. 96(7): 2983-3012

Chan, M. K., Kim, J. and Rees, D. C. 1993. The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 Å resolution structures. Science. 260(5109): 792-4

Chapin, S. F. III, Matson, P. A. and Mooney H. A. 2002. Principles of Terrestrial Ecosystem Ecology, Springer Publishers, New York, NY, USA.

Chisnell, J. R., Premakumar, R. and Bishop, P. E. 1988. Purification of a second alternative nitrogenase from a *nifHDK* deletion strain of *Azotobacter vinelandii*. J. Bacteriol. 170: 27-33

Christiansen, J., Chan, J. M., Seefeldt, L. C., and Dean, D. R. 2000. The role of the MoFe protein □-125^{Phe} and □-125^{Phe} residues in *Azotobacter vinelandii* MoFe protein-Fe protein interaction. J. Inorg. Biochem. 80: 195–204

Christiansen, J., Dean, D. R. and Seefeldt, L. C. 2001. Mechanistic features of the Mo-containing nitrogenase. Ann. Rev. Plant. Physiol. Plant. Mol. Biol.., 52: 269-295.

Christiansen, J., Goodwin, P. J., Lanzilotta, W. N., Seefeldt, L. C. and Dean, D. R. 1998. Catalytic and biophysical properties of a nitrogenase Apo-MoFe protein produced by a *nif*B-deletion mutant of *Azotobacter vinelandii*. *Biochemistry* 37: 12611–12623

Corbett, M. C., Hu, Y., Fay, A. W., Ribbe, M. W., Hedman, B. and Hodgson, K. O. 2006. Structural insights into a protein-bound iron-molybdenum cofactor precursor. Proc. Natl. Acad. Sci. USA 103(5): 1238-1243

Cordewener, J., Kruse-Wolters, M., Wassink, H., Haaker, H., and Veeger, C. 1988. The role of MgATP hydrolysis in nitrogenase catalysis. Eur. J. Biochem. 172: 739–745

Dance, I. 2006. Mechanistic significance of the preparatory migration of hydrogen atoms around the FeMoco active site of nitrogenase. Biochemistry. 45(20): 6328-6340.

Dance, I. 1996. Theoretical investigations of the mechanism of biological nitrogen fixation at the FeMo cluster site. J. Biol. Inorg. Chem. 1: 581–586

Dixon, R. and Daniel, K. 2004. Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol. 2: 621-631

Dos Santos, P. C., Dean, D. R., Hu, Y. and Ribbe, M. W. 2004. Formation and insertion of the nitrogenase iron-molybdenum cofactor. Chem. Rev. 104(2): 1159-1173.

Dos Santos, P. C., Smith, A. D., Frazzon, J., Cash, V. L., Johnson, M. K., Dean, D. R. 2004. Iron-sulfur cluster assembly - NifU directed activation of the nitrogenase Fe protein. J. Biol. Chem. 279(19):19705-19711

Dyer, D. H., Rubio, L. M., Thoden, J. B., Holden, H. M., Ludden, P. W. and Rayment, I. 2003. The three-dimensional structure of the core domain of NafY from *Azotobacter vinelandii* determined at 1.8-A resolution. 278(34): 32150-6.

Eady, R. R. 1988. The vanadium-containing nitrogenase of *Azotobacter*. Biofactors 1(2): 111-116.

Eady, R. R. 1992. The dinitrogen fixing bacteria. In: A. Balows, H.G. Trüper, M. Dworkin, W. Harder and K.H. Schliefer, (2nd edit. ed.), The Prokaryotes, pp. 534–553, Springer-Verlag, New York, NY, USA

Eady, R. R. 1996. Structure function relationships of alternative nitrogenases. Chem. Rev. 96(7): 3013-3030

Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. and Rees, D. C. 2002. Nitrogenase MoFe-protein at 1.16 Å resolution: A central ligand in the FeMo-cofactor. Science. 297(5587): 1696-1700

Furley, P.A. and Newey, W. W. 1983. Geography of the Biosphere. Butterworths, London, UK.

Galloway, J. N., Schlesinger, W. H., Levy, H., Muchat, A. and Schnoor, J.L. 1995. Nitrogen fixation: Anthropogenic enhancement-environmental response. Global Biogeochem. Cycles. 9: 235-252.

Gavini, N. and Pulakat, L. 2002. Role of NifM in maturation of the Fe protein of nitrogenase *In*: T. M. Finan, M. R. O'Brian, D. B. Layzell, J. K. Vessey and W. Newton (Ed.), Nitrogen Fixation - global perspectives. CABI Publishing, New York, NY

Gavini, N., Tungtur, S. and Pulakat, L. 2006. Peptidyl-prolyl cis/trans isomerase-independent functional NifH mutant of *Azotobacter vinelandii*. J. Bacteriol. 188(16): 6020-6025.

Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc J. J. and Rees, D.C. 1992. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. Science. 257(5077): 1653-9.

Goodwin, P. J., Agar, J. N., Roll, J. T., Roberts, G. P., Johnson, M. K. and Dean, D. R. 1998. The *Azotobacter vinelandii* NifEN complex contains two identical [4Fe-4S] clusters. Biochemistry. 37(29): 10420-10428

Hageman, R. V. and Burris, R. H. 1978. Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle. Proc. Natl. Acad. Sci. USA 75(6): 2699-2702

Hawkes, T. R., McLean, P. A and Smith, B. E. 1984. Nitrogenase from *nifV* mutants of *Klebsiella pneumoniae* contains an altered form of the ironmolybdenum cofactor. Biochem. J. 217: 317-321

Hill, S., Kennedy, C. and Kavanagh, E. 1981. Nitrogen fixation gene (*nifL*) involved in oxygen regulation of nitrogenase synthesis in *K. pneumoniae*. Nature, 290, 424–426

Hinneman, B. and Norskov, J. K. 2003. Modeling a central ligand in the nitrogenase FeMo cofactor. J. Am. Chem. Soc. 125(6): 1466-1467

Hofmann-Findeklee, C., Gadkari, D. and Meyer, O., 2000. Superoxide-dependent nitrogen fixation. In: Nitrogen Fixation- From Molecules to Crop Productivity, pp.23-30. Kluwer, Boston, MA, USA

Homer, M. J., Paustian, T. D., Shah, V. K. and Roberts, G. P. 1993. The *nifY* product of *Klebsiella pneumoniae* is associated with apodinitrogenase and dissociates upon activation with the iron- molybdenum cofactor.

Hoover, T. R., Robertson, A. D., Cerny, R. L., Hayes, R. N., Imperial, J., Shah, V. K. and Ludden, P. W. 1987. Identification of the V factor needed for synthesis of the iron-molybdenum cofactor of nitrogenase as homocitrate. Nature. 329:855-857

Howard, J. B. and Rees, D. C. 1994. Nitrogenase: a nucleotide-dependent molecular switch. Annu. Rev. Biochem. 63: 235-64

Howard, J. B. and Rees, D. C., 1996. Structural basis of biological nitrogen fixation. Chem. Rev. 96(7): 2965-2982.

Hu, J. C., Kornacker, M. G. and Hochschild, A. 2000. *Escherichia coli* one and two-hybrid systems for the analysis and identification of protein-protein interactions. Methods. 20(1): 80-94

Hu, Y., Fay, A. W. and Ribbe, M. W. 2005. Identification of a nitrogenase FeMo cofactor precursor on NifEN complex. Proc. Natl. Acad. Sci. 102(9): 3236-3241

Igarashi, R. Y. and Seefeldt, L. C. 2003. Nitrogen fixation. The mechanism of the Mo-dependent nitrogenase. Crit. Rev. Biochem. Mol. Biol. 38(4): 351-384

Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E. and Dean, D. R. 1989a. Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*. J. Bacteriol. 171(2): 1017–1027.

Jacobson, M. R., Cash, V. L., Weiss, M. C., Laird, N. F., Newton, W. E. and Dean, D. R. 1989b. Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii.* Mol. Gen. Genet. 219(1-2): 49-57

Jang, S. B., Seefeldt, L. C. and Peters, J. W. 2000. Insights into nucleotide signal transduction in nitrogenase: Structure of an iron protein with MgADP bound. Biochemistry. 39: 14745-14752.

Joerger, R. D. and Bishop, P. E. 1988. Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. J. Bacteriol. 170(4): 1475-87

Kim, J. S. and Rees, D. C., 1992a. Crystallographic structure and functional implications of the nitrogenase molybdenum iron protein from *Azotobacter vinelandii*. Nature, 360(6404): 553-560.

Kim, J. S. and Rees, D. C., 1992b. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. Science, 257(5077): 1677-1682.

Kurnikov, I. V., Charnley, A. K., and Beratan, D. N. 2001. From ATP to electron transfer: electrostatics and free-energy transduction in nitrogenase. J. Phys. Chem. B 105: 5359–5367

Lanzilotta, W. N., Fisher, K. and Seefeldt, L. C. 1996. Evidence for electron transfer from the nitrogenase iron protein to the molybdenum-iron protein without MgATP hydrolysis: characterization of a tight protein-protein complex. Biochemistry 35(22): 7188 – 7196

Lee, S. H., Pulakat, L., Parker, K. C. and Gavini, N. 1998. Genetic analysis on the NifW by utilizing the yeast two-hybrid system revealed that the NifW of *Azotobacter vinelandii* interacts with the NifZ to form higher-order complexes. Biochem. Biophys. Res. Commun. 244(2): 498-504

Lindahl, P. A., Day, E. P., Kent, T. A., Orme-Johnson, W. H., and Münck, E. 1985. Mössbauer, EPR, and magnetization studies of the *Azotobacter vinelandii* Fe protein. Evidence for a $[4Fe–4S]^{1+}$ cluster with spin $S = 3/2$. J. Biol. Chem. 260: 11160–11173

Lovell, T., Liu, T., Case, D. A. and Noodleman, L. 2003. Structural, spectroscopic and redox consequences of a central ligand in the FeMoco of nitrogenase: a density functional theoretical study. 125(27): 8377-8383

Ludden, P., Roberts, G., Lowery, R., Fitzmaurice, W., Saari, L., Lehman, D., Lies, D., Woehle, D., Wirt, H., Murrell, S., Pope, M. and Kanemoto, R. 1988. Regulation of nitrogenase activity by reversible ADP-ribosylation of dinitrogenase reductase *In*:

H. Bothe, F. J., de Brujin and W. E. Newton (ed.), Nitrogen fixation: hundred years after, pp. 157-162. Gustav Fischer, New York, NY, USA.

Ma, L., Brosius, M. A., and Burgess, B. K. 1996. Construction of a form of the MoFe protein of nitrogenase that accepts electrons from the Fe protein but does not reduce substrate. J. Biol. Chem. 271: 10528–10532

Masepohl B, Krey R, Klipp W. 1993. The *draTG* gene region of *Rhodobacter capsulatus* is required for post-translational regulation of both the molybdenum and the alternative nitrogenase. J. Gen. Microbiol. 139(11): 2667-2675

Merrick, M., Hill, S., Hennecke, H., Hanh, M., Dixon, R. and Kennedy, C. 1982. Repressor properties of the *nifL* gene product in *Klebsiella pneumoniae*. Mol. Gen. Genet. 185: 75–81.

Merrick, M. J. and Edwards, R. A. 1995. Nitrogen control in bacteria. Microbiol. Rev. 59(4): 604-622

Moisander, P. H., Shiue, L., Steward, G. F., Jenkins, B. D., Bebout, B. M., Zehr, J. P. 2006. Application of a *nifH* oligonucleotide microarray for profiling diversity of N-fixing microorganisms in marine microbial mats. Environ. Microbiol. 8(10): 1721-1735

Mortenson, L. E., Seefeldt, L. C., Morgan, T. V. and Bolin, J. T. 1993. The role of metal clusters and MgATP in nitrogenase catalysis. Adv. Enzymol. Relat. Areas. Mol. Biol. 67: 299-374

Nishibayashi, Y., Saito, M., Uemura, S., Takekuma, S., Takekuma, H. and Yoshida, Z. 2004. Buckminsterfullerenes: a non-metal system for nitrogen fixation. Nature 428(6980): 279-280.

Pau, R. N., Mitchenall, L. A. and Robson, R. A. 1989. Genetic evidence for an *Azotobacter vinelandii* nitrogenase lacking molybdenum and vanadium. J. Bacteriol. 171: 124-129

Peters, J. W., Fisher, K., Dean, D. R., 1995. Nitrogenase structure and function - a biochemical-genetic perspective. Ann. Rev. Microbiol. 49: 335-366.

Poole, R. K. and Hill, S. 1997. Respiratory protection of nitrogenase activity in *A. vinelandii*--roles of the terminal oxidases. Bioscience. Rep. 17(3): 303-317

Postgate, J. 1998. Nitrogen Fixation, 3rd Edition. Cambridge University Press, Cambridge, UK.

Pratt, J. M. 1978. The chemistry and biochemistry of nitrogen. Horiz. Biochem. Biophys. 5: 119-160.

Raja, K., Pulakat, L., Gavini, N. 2006. Genetic Complementation Studies of Human Pin1 in *Azotobacter vinelandii* Revealed that it Requires Amino Terminus of the NifM to deliver PPIase Effect to the Fe-protein of Nitrogenase. Am. J. Biochem. Biotech. 2(1): 25-32.

Rangaraj, P., Ruttimann-Johnson, C., Shah, V. K., and Ludden, P. W. 2001. Accumulation of ^{55}Fe-labeled precursors of the iron-molybdenum cofactor of nitrogenase on NifH and NifX of *Azotobacter vinelandii*. J. Biol. Chem. 276 (19): 15968-15974

Rediers, H., Vanderleyden, J. and De Mot, R. 2004. *Azotobacter vinelandii*: a *Pseudomonas* in disguise? Microbiology 150(5): 1117-1119

Rees, D.C., Chan, M.K. and Kim, J. 1993. Structure and function of nitrogenase. Adv. Inorg. Chem., 40: 89-119.

Ribbe, M., Gadkari, D. and Meyer, O., 1997. N_2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N_2 reduction to the oxidation of superoxide produced from O_2 by a molybdenum-CO dehydrogenase. J. Biol. Chem. 272(42): 26627-26633.

Ribbe, M.W., Hu, Y., Guo, M., Schmid, B., and Burgess, B.K. 2002. The FeMoco-deficient MoFe protein produced by a *nif* H deletion strain of *Azotobacter vinelandii* shows unusual P-cluster features. J. Biol. Chem. 277**:** 23469–23476.

Roberts, G. P., MacNeil, T., MacNeil, D. and Brill, W. J. 1978. Regulation and characterization of protein products coded by the *nif* (Nitrogen Fixation) genes of *Klebsiella pneumoniae*. J. Bacteriol. 136: 267-279.

Robinson, A. C., Burgess, B. K. and Dean, D. R. 1986. Activity, reconstitution, and accumulation of nitrogenase components in *Azotobacter vinelandii* mutant strains containing defined deletions within the nitrogenase structural gene cluster. J. Bacteriol. 166(1): 180-186

Rodriguez-Quinones, F., Bosch, R. and Imperial, J. 1993. Expression of the *nifBfdxNnifOQ* region of *Azotobacter vinelandii* and its role in nitrogenase activity. J. Bacteriol. 175(10): 2926-2935.

Rubio, L. M., Ludden, P. W. 2005. Maturation of nitrogenase: a biochemical puzzle. J. Bacteriol. 187(2): 405-414

Rubio, L. M., Rangaraj, P., Homer, M. J., Roberts, G. P. & Ludden, P. W. 2002. Cloning and mutational analysis of the ã gene from *Azotobacter vinelandii* defines a new family of proteins capable of metallocluster binding and protein stabilization. J. Biol. Chem. 277: 14299-14305.

Rudolf, M. and Kroneck, P. M. 2005. The nitrogen cycle: its biology. 43: 75-103

Ryle, M. J., Lanzilotta, W. N. and Seefeldt, L. C. 1996a. Elucidating the mechanism of nucleotide-dependent changes in the redox potential of the [4Fe–4S] cluster in nitrogenase iron protein: The role of phenylalanine 135. Biochemistry 35: 9424–9434

Ryle, M. J. and Seefeldt, L. C. 1996b. The [4Fe-4S] cluster domain of the nitrogenase iron protein facilitates conformational changes required for the cooperative binding of two nucleotides. Biochemistry. 35(49): 15654-15662

Schindelin, H., Kisker, C., Schlessman, J. L., Howard, J. B. and Rees, D. C. 1997. Structure of ADP x AlF_4^--stabilized nitrogenase complex and its implications for signal transduction. Nature 387: 370–376

Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B. and Rees, D. C. 1998. Conformational variability in structures of the nitrogenase iron proteins from *A. vinelandii* and *C. pasteurianum.* J. Mol. Biol. 280: 669-85

Schmid, B., Einsle, O., Chiu, H. J., Willing, A., Yoshida, M., Howard, J. B. and Rees, D. C. 2002. Biochemical and structural characterization of the cross-linked complex of nitrogenase: comparison to the ADP-AlF4(-)-stabilized structure. Biochemistry. 41(52): 15557-15565

Schmid, B., Ribbe, M., Einsle, O., Yoshida, M., Thomas, L. M., Dean, D. R., Rees, D. C. and Burgess, B. K. 2002. Structure of a cofactor-deficient nitrogenase MoFe protein. Science. 296(5566): 352-356

Seefeldt, L. C., Dance, I. G. and Dean, D. R. 2004. Substrate interactions with nitrogenase: Fe versus Mo. Biochemistry. 43(6): 1401-9.

Sen, S., Krishnakumar, A., McClead, J., Johnson, M. K., Seefeldt, L. C., Szilagyi, R. K. and Peters, J. W. 2006. Insights into the role of nucleotide-dependent conformational change in nitrogenase catalysis: Structural characterization of the nitrogenase Fe protein Leu127 deletion variant with bound MgATP. J. Inorg. Biochem. 100(5-6): 1041-1052

Shah, V. K., Allen, J. R., Spangler, N. J. and Ludden, P. W. 1994. *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. Purification and characterization of NifB cofactor, the product of NifB protein. J. Biol. Chem. 269(2): 1154-1158

Shah, V. K., Rangaraj, P., Chatterjee, R., Allen, R. M., Roll, J. T., Roberts, G. P. and Ludden, P. W. 1999. Requirement of NifX and Other *nif* Proteins for *In Vitro* Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase. J. Bacteriol. 181(9): 2797–2801.

Shantharam, S. and Mattoo, A. 1997. Enhancing biological nitrogen fixation: An appraisal of current alternative technologies for N input into plants. Plant and Soil. 194: 205-216

Simon, H. M., Homer, M. J. and Roberts, G. P. 1996. Perturbation of *nifT* expression in *Klebsiella pneumoniae* has limited effect on nitrogen fixation. J. Bacteriol. 178 (10) 2975-2977.

Smith, B. E., Lowe, D. J., and Bray, R. C. 1973. Studies by electron paramagnetic resonance on the catalytic mechanism of nitrogenase of *Klebsiella pneumoniae*. Biochem. J. 135:331–341

Spee, J. H., Arendsen, A. F., Wassink, H., Marritt, S. J., Hagen, W. R. and Haaker, H. 1998. Redox properties and electron paramagnetic resonance spectroscopy of the transition state complex of *Azotobacter vinelandii* nitrogenase. FEBS Lett. 432 (1-2): 55-58.

Stephens, P. J., Jollie, D. R., and Warshel, A. 1996. Protein control of redox potentials of iron-sulfur proteins. Chem. Rev. 96(7): 2491 – 2514

Stevenson, F. J. 1982. Origin and distribution of nitrogen in soil. Nitrogen in Agricultural Soils, pp. 1-42. F.J. Stevenson (ed.). Agronomy Monograph No. 22. American Society of Agronomy, Madison, WI., USA

Stevenson, F. J. 1986. Cycles of soil: carbon, nitrogen, phosphorus, micronutrients. Wiley Interscience, New York, NY, USA

Stiefel, E. I. 1996. Molybdenum bolsters the bioinorganic brigade. Science. 272: 1599-1600.

Suh, M. H., Pulakat, L. and Gavini, N. 2003. Functional expression of a fusion-dimeric MoFe protein of nitrogenase in *Azotobacter vinelandii*. J. Biol. Chem. 278(7): 5353-5360

Thorneley, R. N. and Ashby, G. A. 1989. Oxidation of nitrogenase iron protein by dioxygen without inactivation could contribute to high respiration rates of *Azotobacter* species and facilitate nitrogen fixation in other aerobic environments. Biochem. J. 261(1): 181-187

Thorneley, R. N. and Lowe, D. J. 1983. Nitrogenase of *Klebsiella pneumoniae*. Kinetics of the dissociation of oxidized iron protein from molybdenum-iron protein: identification of the rate-limiting step for substrate reduction. Biochem. J. 215(2): 393-403

Tittsworth, R. C. and Hales, B. J. 1993. Detection of EPR signals assigned to the 1-equivalence-oxidized P-cluster of the nitrogenase MoFe-protein from *Azotobacter vinelandii*. J. Am. Chem. Soc. 115: 9763–9767

Ugalde, R. A., Imperial, J., Shah, V. K. and Brill, W. J. 1984. Biosynthesis of iron-molybdenum cofactor in the absence of nitrogenase. J. Bacteriol. 159(3): 888-893

Watt, G. D., Wang, Z. C. and Knotts, R. R. 1986. Redox reactions of and nucleotide binding to the iron protein of *Azotobacter vinelandii*. Biochemistry. 25(25): 8156-8162

Watt, G. D. and Reddy, K. R. N. 1994. Formation of an all ferrous Fe_4S_4 cluster in the iron protein component of *Azotobacter vinelandii* nitrogenase. J. Inorg. Biochem. 53: 281–294.

Weston, M. F., Kotake, S. and Davis, L. C. 1983. Interaction of nitrogenase with nucleotide analogs of ATP and ADP and the effect of metal ions on ADP inhibition. Arch. Biochem. Biophys. 225: 809–817.

Willing, A. and Howard, J. B. 1990. Cross-linking site in *Azotobacter vinelandii* complex. J. Biol. Chem. 265: 6596–6599.

Wilson, P. E., Nyborg, A. C. and Watt, G. D. 2001. Duplication and extension of the Thorneley and Lowe kinetic model for *Klebsiella pneumoniae* nitrogenase catalysis using a MATHEMATICA software platform. Biophys. Chem. 91: 281–304

Zheng, L., White, R. H., Cash, V. L., Jack, R. F. and Dean, D. R. 1993. Cysteine desulfurase activity indicates a role for NifS in metalloclusters biosynthesis. Proc. Natl. Acad. Sci. USA. 90(7): 2754-2758

Zheng, L., White, R. H., Cash, V. L. and Dean, D. R. 1994. Mechanism for the desulfurization of L-cysteine catalyzed by the *nifS* gene product. Biochemistry. 33(15): 4714-4720

Zimmermann, R., Münck, E., Brill, W. J., Shah, V. K., Henzl, M. T., Rawlings, J. and Orme-Johnson, W. H. 1978. Nitrogenase X: Mössbauer and EPR studies on reversibly oxidized MoFe protein from *Azotobacter vinelandii* OP. Nature of the iron centers. Biochem. Biophys. Acta. 537: 185–207

Zumft, W. G., Mortenson, L. E., Palmer, G. 1974. Electron-paramagnetic-resonance studies on nitrogenase. Investigation of the oxidation-reduction behaviour of azoferredoxin and molybdoferredoxin with potentiometric and rapid-freeze techniques. Eur. J. Biochem. 46(3): 525-535

CHAPTER II

FUNCTIONAL NIFD-K FUSION PROTEIN IN *AZOTOBACTER VINELANDII* IS A HOMODIMERIC COMPLEX EQUIVALENT TO THE NATIVE HETEROTETRAMERIC MOFE PROTEIN

(Published in Biochem. Biophys. Res. Commun., 2005; 337(2): 677-84)

Abstract

The MoFe protein of the complex metalloenzyme nitrogenase folds as a heterotetramer containing two copies each of the homologous alpha and beta subunits, encoded by the *nifD* and the *nifK* genes respectively. Recently, the functional expression of a fusion NifD-K protein of nitrogenase was demonstrated in A. vinelandii, strongly implying that the MoFe protein is flexible as it could accommodate major structural changes, yet remain functional (Suh, M.H., Pulakat, L. and Gavini, N. 2003. J. Biol. Chem., 278: 5353-5360). This finding led us to further explore the type of interaction between the fused MoFe protein units. We aimed to determine whether an interaction exists between the two fusion MoFe proteins to form a homodimer that is equivalent to native heterotetrameric MoFe protein. Using the Bacteriomatch™ Two-Hybrid System, translationally fused constructs of NifD-K(fusion) with the full-length □CI of the pBT bait vector and also NifD-K(fusion) with the □□terminal □-RNAP of the pTRG target vector were made. To compare the extent of interaction between the fused NifD-K

proteins to that of the □□□ interactions in the native MoFe protein, we proceeded to generate translationally fused constructs of NifK with the □-RNAP of the pTRG vector and □CI protein of the pBT vector. The strength of the interaction between the proteins in study was determined by measuring the □-galactosidase activity and extent of ampicillin resistance of the colonies expressing these proteins. This analysis demonstrated that direct protein-protein interaction exists between NifD-K fusion proteins, suggesting that they exist as homodimers. As the interaction takes place at the □□interfaces of the NifD-K fusion proteins, we propose that these homodimers of NifD-K fusion protein may function in a similar manner as that of the heterotetrameric native MoFe protein. The observation that the extent of protein-protein interaction between the □-subunits of the native MoFe protein in BacterioMatch Two-Hybrid System is comparable to the extent of protein-protein interaction observed between the NifD-K fusion proteins in the same system further supports this idea.

Introduction

The abundant but inert N_2 in the atmosphere is converted to the metabolically useful NH_3 by the metalloenzyme nitrogenase. Although nitrogen fixation is a property of a phylogenetically diverse set of bacteria and cyanobacteria, in general, the sequences, structures, and functional properties of the nitrogenase are highly conserved between different organisms (Rees, 1993; Peters et al., 1995; Howard and Rees, 1996). Four classes of nitrogenases have been characterized (Stiefel, 1996; Hofmann-Findeklee, 2000). Three of these classes share many similarities, differing in part, by the heterometal atom contained in the active site metal cluster (Mo, V, Fe) (Loveless et al., 1999; Loveless and Bishop, 1999). The fourth class is a superoxide-dependent nitrogenase isolated from *Streptomyces thermoautotrophicus* (Ribbe et al., 1997), which is distinct from the other nitrogenase classes. The Mo-dependent nitrogenases are the best studied and most widely distributed (Burgess and Lowe, 1996; Rajagopalan and Johnson, 1992; Christiansen et al., 2001). Much of the understanding of the structural properties of the Mo-dependent nitrogenase was obtained through studies on the crystallographic structures of both component proteins, the Fe protein and the MoFe protein of nitrogenase and their metal clusters from *Azotobacter vinelandii*, *Klebsiella pneumoniae* and *Clostridium pasterurianum* (Kim and Rees, 1992a; Kim and Rees, 1992b; Georgiadis et al., 1992; Strange et al., 2003; Mayer et al., 2002; Schlessman et al., 1998; Kim et al., 1993). The Fe protein has a molecular weight of about 60 kDa and is a dimer of identical subunits encoded by the nifH gene. Both subunits are bridged by one 4Fe-4S metal

center and contain two nucleotide [MgATP or MgADP] binding sites (Kim and Rees, 1994; Georgiadis et al., 1992; Jang et al., 2000; Moshiri et al., 1995; Peters et al., 1995; Schindelin et al., 1997). The 230 kDa MoFe protein is a tetramer in its biologically active form and is composed of two identical halves, each containing an □ subunit and a □ subunit encoded by the *nifD* and *nifK* genes, respectively (Howard and Rees, 1996; Kim and Rees, 1992a; Schindelin et al., 1997; Burgess and Lowe, 1996). The □ and □ subunits of Azotobacter vinelandii MoFe protein have 491 and 522 amino acids respectively (Brigle et al., 1985). The MoFe protein contains two types of metal centers, the FeMo-cofactor, the substrate reduction site; and the P-cluster, which participates in electron transfer from the Fe protein to the FeMo-cofactor (Howard and Rees, 1996; Schindelin et al., 1997; Burgess and Lowe, 1996; Peters et al., 1997). Although there is minimal amino acid sequence homology between the □ and □ subunits of the A. vinelandii MoFe protein, they are known to exhibit similar polypeptide folds, which consist of three domains of the parallel □-sheet / □-helical type (Kim and Rees, 1992a; Kim and Rees, 1992b). At the interface between the three domains is a wide, shallow cleft; in the □ subunit, the FeMo cofactor occupies the bottom of this cleft. The P cluster is buried at the interface between a pair of □ and □ subunits with a pseudo-2-fold rotation axis passing between the two halves of the P cluster and relating the two subunits. The extensive interaction between □ and □ subunits in an □□ dimer suggests that they form the fundamental functional unit. An open channel of ~8 Å diameter exists between the two pairs of □□ dimers with the tetramer 2-fold axis extending through the

center. The tetramer interface is dominated by interactions between helices from the two □ subunits, along with a cation binding site, presumably occupied by calcium, that is coordinated by residues from both □ subunits (Howard and Rees, 1996; Blanchard and Hales, 1996).

Suh, M. H. et al. have recently shown that the MoFe protein could be functional when synthesized as a single protein encoded by a *nifD-K* translational fusion in A. vinelandii. The fused NifD-K protein encoded by the *nifD-K* translational fusion had a total loss of three amino acids and seven mismatches from the wild type NifD and NifK proteins. Even after such an alteration, the NifD-K fused protein was found to be capable of supporting nitrogen fixation when expressed in *A. vinelandii* (Suh et al., 2003), implying that the MoFe protein is flexible and can accommodate major structural changes, yet remain functional. In the wild type MoFe protein encoded by *A. vinelandii*, the NifD and NifK are synthesized separately and then assembled together. However, the NifD-K fusion strain, *A. vinelandii* BG1304 showed a significant difference in the biosynthesis and assembly pattern of its MoFe protein than that of the wild type MoFe protein, as it formed a single large NifD-K fusion protein (Suh et al., 2003). Based on the above result of obtaining a functional NifD-K fusion protein in *A. vinelandii*, we aimed to further explore the type of interaction existing amongst the fused NifD-K proteins, mainly to test whether the NifD-K fusion protein functions as a single unit or a dimer. We also compared the extent of interaction between the NifD-K fusion protein with that of the □-subunits of the native MoFe protein in order to understand the nature of the higher order complex of the MoFe protein. We also investigated whether direct or

physical protein-protein interaction is involved in the activity of the functional MoFe protein in *A. vinelandii* BG1304 and which subunits are particularly involved in such an interaction.

The BacterioMatch™ Two-Hybrid System (Dove et al., 1997; Dove and Hochschild, 1998) is a recently developed molecular genetic approach for detection of protein-protein interactions, *in vivo* in *E.coli*, based on the principle of transcriptional activation. A protein of interest (the bait) is fused to the full-length bacteriophage □CI protein, containing the amino-terminal DNA-binding domain and the carboxyl terminal dimerization domain. The corresponding target protein is fused to the N-terminal domain of the □ subunit of RNA polymerase. The bait is tethered to the □ operator sequence upstream of the reporter promoter through the DNA-binding domain of □CI. When the bait and target interact, they recruit and stabilize the binding of RNA polymerase to the promoter and activate transcription of the *ampR* reporter gene. A second reporter gene, lacZ is expressed from the same promoter, providing additional validation of the bait and target interaction. This unique *E. coli* based system allows many significant advantages over other two hybrid systems since it offers faster results and detects an interaction between a pair of protein domains with an equilibrium dissociation constant in the high nanomolar range (Dove et al., 1997; Stratagene, Inc., 2001). The results reported here by utilizing the two-hybrid system reveal the existence of an interaction among the NifD-K fusion proteins. The evidence provided in this report indicates that the NifD-K fused proteins interact with each other, suggesting that they may exist as higher order dimeric NifD-K fusion proteins that are equivalent to heterotetrameric native MoFe proteins.

Also, our results show that the interacting surfaces include the NifK of one NifD-K protein and the NifK of the other NifD-K fusion protein, implying that the interaction takes place at the □□□ surfaces.

Materials and Methods

Strains, plasmids and growth conditions. The Bacterial strains and plasmids used in this study are described in Table 2.1. *E. coli* strains were normally grown at 37°C in 2YT media (Sambrook et al., 1992). Ampicillin, chloramphenicol, and tetracycline were used to a final concentration of 50, 34, and 5 μg/ml respectively, wherever the selection was made.

General Molecular Techniques. Restriction enzymes were purchased from Promega, Inc. (Madison,WI). DNA sub-cloning, plasmid DNA isolations, restriction enzyme digestions, agarose gel electrophoresis, ligations and *E. coli* transformations were carried out as described in the laboratory manual or according to the manufacturer's protocol (Stratagene, Inc., 2001; Sambrook et al., 1992). Oligonucleotides used for PCR amplification were purchased from GIBCO BRL Life Technologies, Inc (Gaithersburg, MD).

The specific primers used to PCR amplify *nifD-K* fusion from pBG1404 for cloning into the pTRG target vector were: (1) 5'CTC GAG TTA GCG TAC CAG GTC GTG GTT GTA GTC GGT GGC CTG C-3' and (2) 5'CTC GAG ATG ACC GGT ATG TCG CGC GAA GAG GTT GAA TCC CTC ATC-3'. Primers designed to amplify *nifD-K* fusion for cloning into the pBT bait vector, were: (3) 5'CTC GAG TTA GCG

TAC CAG GTC GTG GTT GTA GTC GGT GGC CTG C-3' and (4) 5'GAT ATC TGA CCG GTA TGT CGC GCG AAG AGG TTG AAT CCC TCA TC-3'.

Table 2.1. Bacterial Strains and plasmids used in this study.

Strain/plasmid	Relevant characteristics and description	Reference
Escherichia coli TG1	*K-12 Δ(lac-pro) supE thi hsd-5/F' traD36 proA⁺B⁺ lacI^q lacZΔM15.*	Amersham Life Sciences, Inc.
Escherichia coli XL-1 Blue	*MRF' K, Δ(mcrA) 183Δ[mcr CB-hsdSMR-mrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac (F' proAB lacI^qZ ΔM15 Tn5(Kan)^r]*	Stratagene Corp.
PCR 2.1 TOPO	*Amp^r, Kan^r,* (3.9 kb), used for direct cloning of PCR products, *lacZα* fragment, MCS, M13.	Invitrogen Corp.
pBT	*Cm^R*, 52bp MCS, 3.2 kb size, MCS, p15A origin of replication, *lac-UV5* promoter, □c1 ORF.	Stratagene Corp.
pTRG	*Tet^R*, 60bp MCS, 4.4 kb size, MCS, lac-UV5 promoter,ColE1 origin of replication, RNAP□ ORF.	Stratagene Corp.
pDB6	6 kb *SmaI* fragment spanning *A.vinelandii nifHDKY* genes cloned in the *SmaI* site of pUC8; *Amp^r*, with ColE1 origin of replication.	(Brigle et al., 1985).
pBG1404	Derivative of pUC19 in which a 1510 bp fragment corresponding to *nifD* gene and a 1552 bp fragment corresponding to *nifK* gene were ligated, creating a fused *nifD-K* gene.	(Suh et al., 2003)
pBG2323	Derivative of pTRG in which 1471 bp DNA fragment of *nifD* was cloned to generate an in-frame fusion with □-RNAP-*nifD*. This fragment could be released by digesting with *BamHI*.	(Patil et al., 2003)
pBG1710	Derivative of TOPO PCR 2.1 in which 3074 bp fragment of *nifD-K* fusion gene was cloned directly after PCR. The fragment was PCR amplified using appropriate oligonucleotide primers. It could be released by digesting with *XhoI*.	This work
pBG1711	Derivative of pTRG in which 3072 bp fragment corresponding to *nifD-K* fusion gene from XhoI digested pBG1710 was fused with the □-RNAP of the *Xho*I digested pTRG target vector.	This work
pBG1712	Derivative of TOPO PCR 2.1 in which 3074 bp fragment of *nifD-K* fusion gene was cloned directly after PCR. appropriate oligonucleotide primers designed with *XhoI* and *EcoRV* sites at 5' and 3' ends respectively. It could be released by digesting with *XhoI-EcoRV*.	This work
pBG1713	Derivative of pBT in which 3070 bp fragment to *nifD-K* fusion gene from *XhoI-EcoRV* digested pBG1712 was fused with the nucleotide sequence encoding the full length □CI protein of the *SmaI-XhoI* digested pBT bait vector.	This work
pBG1715	Derivative of TOPO PCR 2.1 in which a specific region (2335 bp) of *nifKTY* was cloned directly after PCR. The fragment was PCR amplified using appropriate oligonucleotide primers designed with *BamHI* sites at both ends. It could be released by digesting with *BamHI*.	This work
pBG1716	Derivative of pTRG in which 2333 bp *nifKTY* fragment released from *BamHI* digested pBG1715 was fused to the *BamHI* digested pTRG.	This work
pBG1718	Derivative of pBT in which 2333 bp *nifKTY* fragment released from *BamHI* digested pBG1715 was fused to the *BamHI* digested pBT.	This work

Also to amplify the *nifK* region, the following primers were used: (5) 5'GGA TCC AAG CGC GAC GGC TTC GAG GAA AAG TAT CCG CAG -3' and (6) 5'GGA TCC CGC GGC AGG GGC CGG GTT CGC GTA GAT CGA TCA -3'. The bacterial two-hybrid vectors, pBT bait and pTRG target, containing the □CI and □-RNAP domains respectively, were the starting plasmids used for cloning *nifD-K* fusion gene. The 3062 bp *nifD-K* fused gene was first PCR amplified from the pBG1404 template using primers (1), (2), (3) and (4) as mentioned above. Primers (1) and (2) were designed to introduce an *Xho*I site in them. The *nifD-K* PCR product was cloned into PCR 2.1 TOPO to generate plasmid pBG1710. The pBG1710 was then digested with *Xho*I and the resulting fragment carrying the *nifD-K* fusion gene was cloned into the *Xho*I digested pTRG vector. Miniscreen and restriction digestion analysis of the subsequent transformants obtained allowed selection of the desired clone, named as pBG1711, containing □-RNAP of the pTRG fused with the *nifD-K* fusion gene. Primers (3) and (4) designed to contain restriction sites *Xho*I and *Eco*RV were used to obtain the *nifD-K* fusion gene for ligating with pBT. First, the PCR amplified fragment was cloned into TOPO 2.1 and named pBG1712. The plasmid pBG1712 was then digested with *Xho*I and *EcoR*V to generate the fragment carrying *nifD-K* fusion gene and having *Xho*I and *EcoR*V ends. This fragment was then cloned into *Xho*I-*Sma*I digested pBT vector to generate the plasmid pBG1713. Since the *SmaI* generated blunt ends, pBT could be ligated with the blunt ended *Eco*RV end of the *Eco*RV-*Xho*I generated *nifD-K* fusion fragment from pBG1712. As shown in Fig. 2.1, subsequently, pBG1711 (□-RNAP + NifD-K fusion) and pBG1713 (full-length □CI + NifD-K fusion) were cotransformed into *E. coli* XLI Blue cells

containing the reporter cassette. The construction of the plasmid pBG2323 carrying □-RNAP of pTRG vector fused to *nifD* is described elsewhere (Patil et al., 2003). The plasmid pBG1715 was generated by inserting the 2335 bp fragment carrying *nifK*, PCR amplified from the template pDB6 (Brigle et al., 1985), using primers (5) and (6), into PCR 2.1 TOPO. The pTRG and pBT plasmids were digested with *Bam*HI and the *Bam*HI fragment obtained from pBG1715 was cloned into both of these vectors, resulting into plasmids pBG1716 and pBG1718 respectively. The plasmids pBG1716 (□-RNAP+NifK) and pBG1718 (full-length □CI +NifK) were cotransformed into *E. coli* XL1 Blue cells.

□-*galactosidase assay*. The □-galactosidase activity assay was performed as described in 'Molecular Cloning' (Sambrook et al., 1992). Briefly, single *E. coli* transformants were inoculated into 5 ml of 2YT media supplemented with 34 □g/ml of chloramphenicol and 5□g/ml of tetracycline. The cells were then incubated overnight at 37°C with shaking at 250 rpm. Then 200□l of the overnight culture was diluted into fresh 5ml of the same media and incubated at 37°C with shaking, until $O.D_{600}$ was between 0.3-0.6 (initial $O.D_{600}$ was 0.05). Cells from 1.5 ml of the culture were collected by centrifugation and resuspended in 500□l of Z-buffer (Sambrook et al., 1992). A 100□l aliquot of the resuspended cells were lysed by adding 50□l of chloroform and 25□l of 0.1% SDS and kept at 28°C with shaking for 5-10 min. To measure the □-galactosidase activity from the cell lysate, 900□l of Z-buffer-□-mercaptoethanol solution (0.27ml of □-mercaptoethanol per 100ml of Z-buffer) was added to 100□l of culture followed by 200□l of Z-

buffer/ONPG (o-nitrophenyl-galactose, 4mg ONPG per ml of Z-buffer). The time of ONPG addition was recorded, and the tubes were incubated at 37°C. When yellow color was visible, 500□l of 1 M Na_2CO_3 was added to each tube to terminate the reaction, and the time was recorded. The optical densities at 420 nm as well as 550 nm were recorded. The □-galactosidase units were defined as the amount of enzyme which hydrolyzes 1 □mole of ONPG to o-nitrophenol and D-galactose per minute. □-galactosidase activity in Miller Units was calculated as $1000 \times [O.D_{420} - (1.75 \times O.D_{550})/ (t \times V \times O.D_{600})]$ where: t = elapsed time (in min) of incubation, V= 0.1 ml x concentration factor, and $O.D_{600}$ = Absorbance of 1 ml of culture at 600 nm. The results of the □-galactosidase assay were verified for significance by using the student's t-test (GraphPad software: http://www.graphpad.com/quickcalcs/ index.cfm)

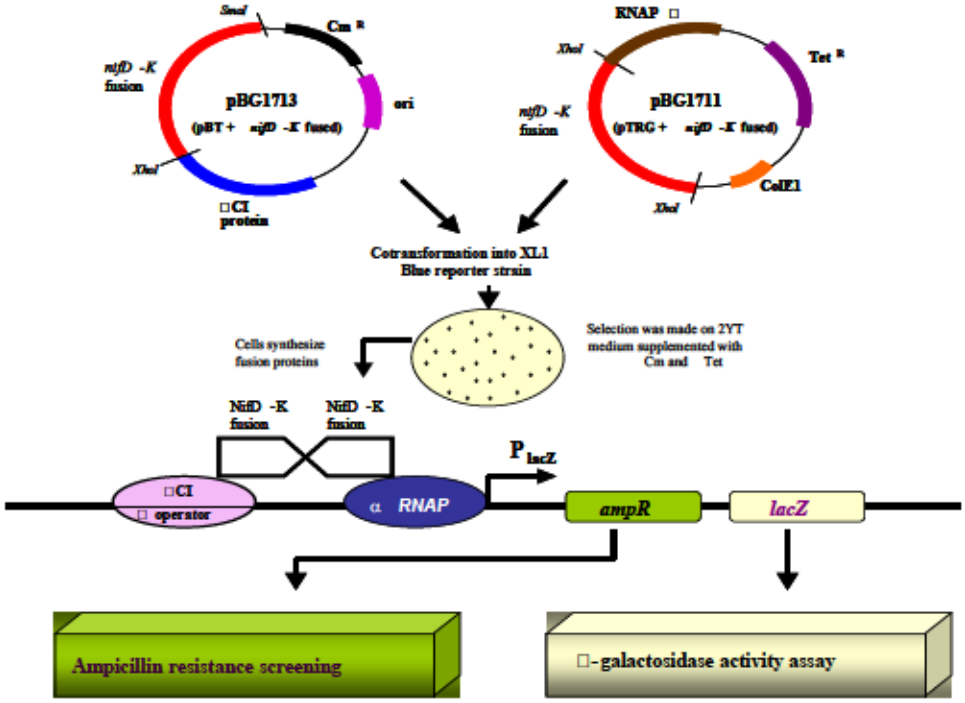
Results and Discussion

Detection of Interaction between NifD-K fusion protein units. Recently, the functional expression of the MoFe protein as a fusion protein was shown in A. vinelandii, wherein the *nifD* and *nifK* genes were translationally fused, thus creating a total loss of three amino acids and seven mismatches, as compared to the wild type NifD and NifK proteins (Suh et al., 2003). It was interesting then, to study the interaction between the NifD-K fusion protein units. To determine presence of any such interaction in between these two fused protein units, we took advantage of the BacterioMatch™ Two-Hybrid protein-protein interaction assay. This strategy was preferred over the conventional yeast based two-hybrid system as it is faster and simpler compared to the latter and also the results

obtained hold a similar degree of sensitivity and reliability. Moreover, it offers the feasibility of studying proteins of prokaryotic origin in a prokaryotic system (Dove et al., 1997; Stratagene, Inc., 2001). The *E.coli* XL1 Blue reporter strain was transformed with the pBT and pTRG vectors of the bacterial two-hybrid system carrying the *nifD-K* fusion gene translationally fused to □CI and □-RNAP respectively (Fig. 2.1 and Fig. 2.2). Upon interaction between the NifD-K fusion proteins fused to the bait and target vectors, the □CI repressor protein attaches to the lambda operator and brings the □-subunit of the RNA polymerase to the promoter to initiate transcription of the *ampR* and *lacZ* reporter genes.

We therefore performed the liquid □-galactosidase assay (Sambrook et al., 1992) for detection of interaction between the NifD-K fusion protein units. The negative controls used in this experiment were generated by cotransforming the following plasmid combinations into the *E.coli* XL1 Blue reporter cells: (a) pBG1711 (N-terminal □-RNAP of pTRG + NifD-K fusion protein) and pBT (full-length □CI only) (b) pBG1713 (full length □CI + NifD-K fusion protein) and pTRG (N-terminal □-RNAP only) and (c) □CI of pBT and □-RNAP of pTRG vectors without any translational fusions. As shown in Table 2.2, the colonies carrying pBG1711 (N-terminal □-RNAP of pTRG fused to NifD-K fusion protein) and pBG1713 (full-length □CI of pBT fused to NifD-K fusion protein) showed a comparatively higher level of □-galactosidase activity, indicating presence of interaction between the NifD-K fusion protein units. The positive control used in this experiment was generated by cotransforming the following plasmids into the *E.coli* XL1

Blue reporter cells; pBT-LGF2 containing the full length □CI protein of pBT vector fused to the LGF2 (dimerization domain of the yeast transcriptional activator Gal4), and pTRG-Gal11^{P} containing the N-terminal domain of the □-RNAP of the pTRG vector fused to the Gal11^{P} (a mutant form of Gal11 protein) (Stratagene, Inc., 2001).



The nifD-K fusion gene was cloned into the *XhoI* site of the pTRG target vector MCS and this plasmid designated as pBG1711. Similarly pBT bait vector was digested with *XhoI* and *SmaI* and the nifD-K fusion gene was cloned into this site. This construct was designated as pBG1713. The plasmids pBG1711 and pBG1713 were used for cotransformation of XL1 Blue reporter *E. coli* cells. An interaction between the NifD-K fused proteins leads to stabilization of the recruitment and binding of the RNA polymerase, thus activating transcription of the downstream reporter genes *lacZ* and *ampR*.

Fig 2.1. Experimental strategy used to detect interaction between NifD-K fusion protein units.

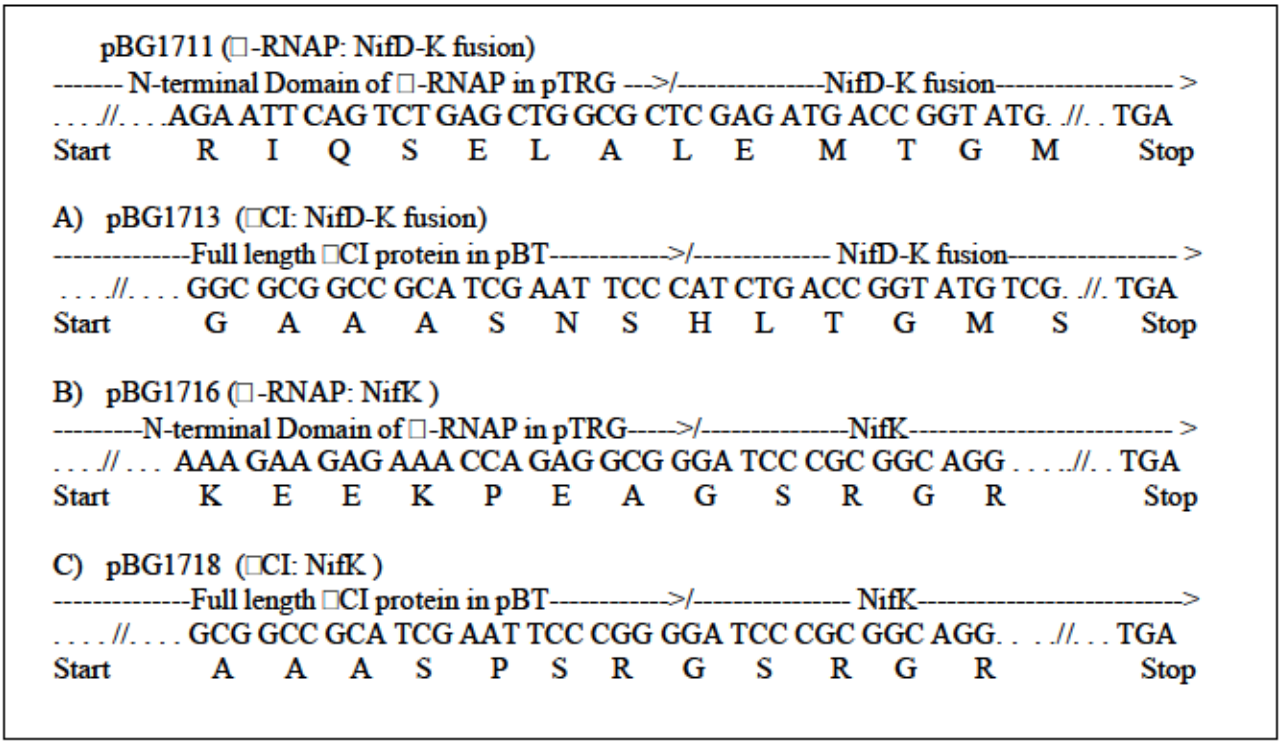
A much higher expression of □-galactosidase was found for the colonies harboring pBT-LGF2 and pTRG-Gal11^{P}, as shown in Table 2. 2. Subsequently, we compared the ampicillin resistance of the control strains and *E.coli* XL1 Blue reporter strain that was cotransformed with pBG1711 (□-RNAP::NifD-K fusion) and pBG1713 (□CI::NifD-K fusion), by serially spotting 10 fold dilutions of the cells on plates containing different concentrations of ampicillin such as 200, 300 and 400 µg/ml. As shown in Fig. 2.3, the colonies expressing NifD-K fusion protein units fused to □-RNAP and □CI showed over 10 fold higher resistance to ampicillin than the negative control carrying pBT and pTRG. Therefore, these results implied that interaction between the NifD-K fusion protein units led to recruitment and stabilization of the binding of RNA polymerase close to the promoter and activation of transcription of the ampR gene. The observation that colonies carrying different combinations of negative control plasmids showed only low basal transcriptional activation of both reporter genes confirmed that the results are not due to any false positives but due to specific protein-protein interactions (Table 2.2).

Interaction between the NifD-K fusion proteins is comparable to the interaction between the □□□ subunits of the native MoFe protein. The detection of interaction between the NifD-K fusion protein units led us to explore which of the subunits, NifD or NifK, was involved in this interaction. We PCR amplified and cloned a 2335 bp fragment encoding the NifK protein spanning amino acids 27 to 523 into the pTRG and pBT vector plasmids, fused to the N-terminal □-RNAP and □CI respectively. Thus, plasmids pBG1716 and pBG1718 were created respectively (Refer 'Materials and Methods').

Table 2.2. Results of the liquid □-galactosidase assay with ONPG as substrate to demonstrate protein-protein interaction.

Plasmid to which □CI was translationally fused	Plasmid to which □-RNAP was translationally fused	□-galactosidase Activity (Miller units)*	Growth on Ampicillin Plates	Interacting Peptides
pBG1713 (□CI:Nif D-K fusion)	pBG1711 (RNAP:NifD-K fusion)	44±2.90	++	NifD-K and NifD-K
pBG1713 (□CI:Nif D-K fusion)	pBG1716 (RNAP:NifK)	79±15.23	++	NifD-K and NifK
pBG1718 (□CI:NifK)	pBG1716 (RNAP:NifK)	48±1.97	++	NifK and NifK
pBG1713 (□CI:Nif D-K fusion)	pBG2323 (RNAP:NifD)	26±3.15	+	None
pBT	pBG1711 (RNAP:NifD-K fusion)	24±1.50	+	None
pBG1713 (□CI:Nif D-K fusion)	pTRG	23±3.80	+	None
pBT	pBG1716 (RNAP :NifK)	35±1.42	+	None
pBT	pTRG	28±7.50	+	None
pBT-LGF2	pTRG-Gal11^{P}	122±36.29	+++	mutant form of Gal1 protein with Gal11P.

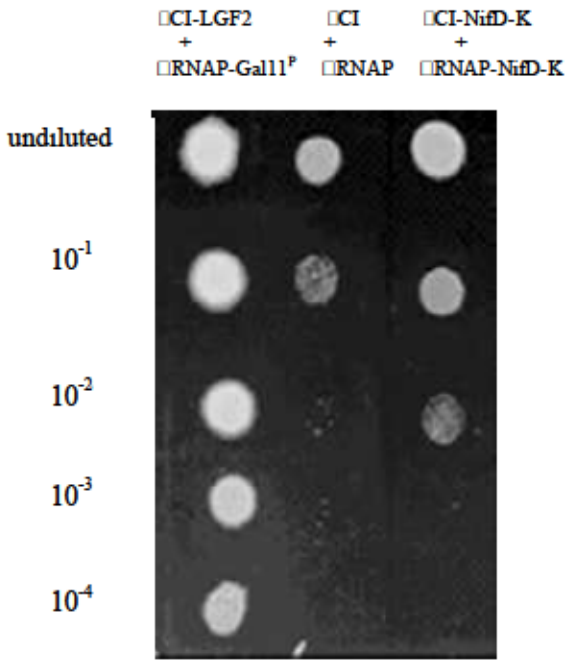
*Each assay was performed a minimum 5 times for accuracy and the □-galactosidase activity units shown are an average of 3 independent observations.



The nucleotide sequences corresponding to NifD-K (fused) and NifK are shown in bold. (A) pBG1711: Fusion junction of the amino terminal domain of the □-RNAP in pTRG and the NifD-K fusion protein. (B) pBG1713: Fusion junction of the full length □CI protein in pBT and the NifD-K fusion protein. (C) pBG1716: Fusion junction of the amino terminal domain of the □-RNAP in pTRG and the NifK sequence. (D) pBG1718: Fusion junction of the full length □CI protein and the NifK sequence.

Fig. 2.2. The nucleotide sequence and respective amino acid designations of junctions of translational fusions in pBG1711, pBG1713, pBG1716 and pBG1718.

For measuring the □-galactosidase activity to detect interaction between the NifK proteins, we cotransformed the plasmids pBG1716 and pBG1718 into *E.coli* XL1 Blue reporter cells. The extent of increase in the □-galactosidase activity of the colonies harboring pBG1716 and pBG1718 when compared to the colonies harboring pBT and pTRG vectors (negative control) were supportive of the concept that the interaction



The O.D_{600} of the positive control (harboring known interacting peptides $Gal11^P$ and LGF2 fused to amino terminal domain of □-RNAP in pTRG and the full-length □CI protein in pBT respectively), negative control (harboring the pBT and pTRG vectors), and E.coli XL1 Blue reporter cells (harboring the experimental interacting NifD-K fusion protein units fused to the N-terminal domain of □-RNAP in pTRG and the full-length □CI in pBT respectively) was 0.3. This was subjected to serial dilutions (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}) and 2µlof the undiluted and subsequent dilutions were then spotted on a 2YT plate supplemented with ampicillin. Growth can be observed till higher dilutions when the ampR gene is transcribed due to recruitment of two positively interacting proteins. The positive colony shows comparatively high ampicillin resistance indicated by growth till 10000 fold dilution. The XL1 Blue cells containing pBG1711 and pBG1713 show growth till 100 fold dilution whereas the negative control shows meagre growth from the 10 fold dilution itself.

Fig. 2.3. Detection of interaction between NifD-K fusion protein units by observation of growth on 2YT plate supplemented with 300 mM ampicillin.

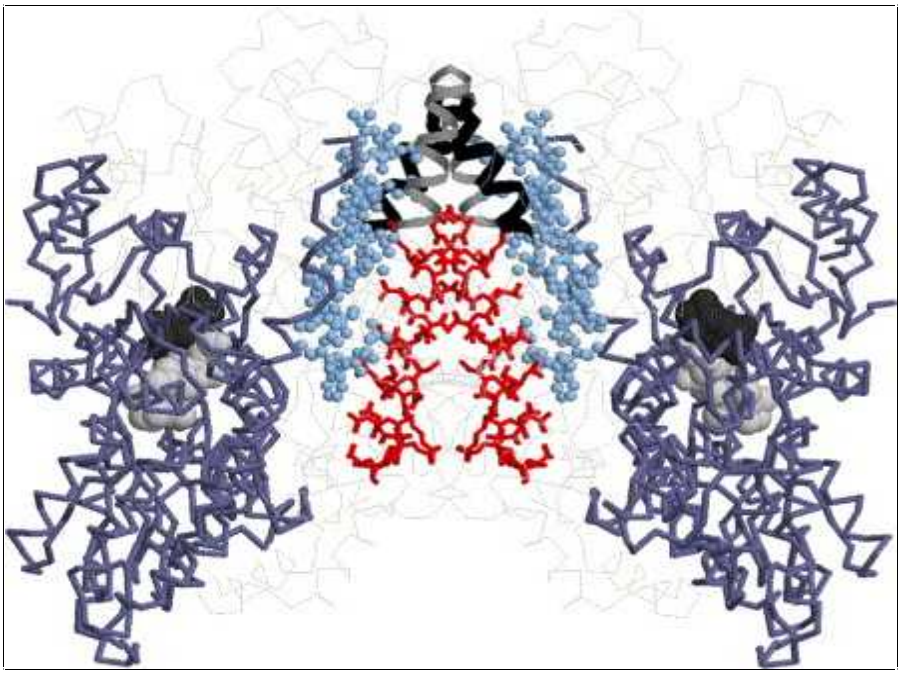
between the subunits of the MoFe protein is primarily due to the interaction between the □-subunits (NifK-NifK interaction). Also, the □-galactosidase activity of the colonies harboring pBG1711 (□-RNAP::NifD-K) and pBG1713 (□CI::NifD-K) were found to be

similar to that of the colonies harboring pBG1716 and pBG1718 (Table 2.2), thus indicating that the strength of interaction between the NifK and NifK proteins is comparable to that of the NifD-K fusion protein units. We analyzed the ampicillin resistance of the strains harboring the pBT and pTRG plasmids, pBT-LGF2 and pTRG-Gal11^{P} plasmids, and the pBG1716 and pBG1718 plasmids by serially spotting 10 fold dilutions of the cells on plates containing different concentrations of ampicillin such as 200, 300 and 400 µg/ml (data not shown). Growth observed was comparable to that found for the colonies harboring pBG1711 and pBG1713, again strongly indicating that the strength of interaction between the NifD-K fusion protein units is comparable to that between the native □-subunits of the MoFe protein. We further checked for the interaction between the NifD-K (fused) and NifK proteins by cotransforming the plasmids pBG1713 (□CI::NifD-K) and pBG1716 (□-RNAP::NifK) into the XL1 Blue reporter cells. These colonies showed almost 3 times higher □-galactosidase activity when compared to the colonies harboring the pBT and pTRG (negative control) plasmids. We also obtained approximately 1.7 fold higher □-galactosidase activity corresponding to the interaction between NifD-K fusion and NifK proteins, compared to that of the NifD-K fusion protein units and that of the NifK proteins (Table 2.2). It maybe that there was an improved structural folding of NifK in such a specific kind of interaction and thus better chances of recruitment and stabilization of the RNA polymerase involved in transcription of the downstream lacZ reporter gene. We confirmed that the NifD does not participate in the interaction between the NifD-K fusion protein units by determining the □-galactosidase activity of the colonies harboring pBG1713 (full-length □CI protein fused

to NifD-K fusion) and pBG2323 (N-terminal □-RNAP fused to NifD). It was found that colonies containing pBG1713 and pBG2323 showed □-galactosidase activity as little as that of the colonies harboring pBT and pTRG (negative control) plasmids, thereby indicating that NifD does not participate in the interaction between the NifD-K fused protein units. When compared to the □-galactosidase activity corresponding to the interaction between NifD-K (fused) and NifK proteins, 3 fold less activity was found for the colonies containing pBG1713 and pBG2323, confirming that only NifK (and not NifD) has a role to play in the interaction between the NifD-K fused protein units.

The X-ray crystallographic structure of the MoFe protein determined by the Rees group (Kim and Rees, 1992a; Kim and Rees, 1992b) forms much of the basis of our knowledge of this intriguing protein at present. According to the X-ray studies, the two □-subunits of the MoFe protein form almost exclusive regions of contact between the □□ pairs. Structurally therefore, the □-□ interface is thought to be generated by packing between helices comprised of residues 234-247 (Domain I), 323-336 (Domain II), 342-362 (Domain III) and 488-510 (Domain IV) (Fig. 2.4).

The construction of a fused □□ (NifD-K fusion) protein (Suh et al., 2003) enabled us to study the interactions between a pair of □□ units jointly. Since our results demonstrate direct protein-protein interaction between the NifD-K fusion protein units, we suggest that they exist as homodimers. Also, a comparison of the strength of interaction between the fused NifD-K proteins to that of the □-□ interactions in the native MoFe protein (as shown by the □-galactosidase activity of the colonies expressing the



NifK (â-subunits) represented in light gray thin backbone structure; NifD (á-subunits) shown in dark gray thick backbone; The â-â interface of the MoFe protein generated by the following domains in both â-subunits is shown as: I—residues 234–247 (black ribbons), II—residues 323–336 (gray ribbons), III—residues 342–362 (dotted spacefill), IV—residues 488–510 (sticks); the larger spacefilled structures represent the FeMo-cofactor (light gray) and the P-Cluster (black). (Kim and Rees, 1992a; Kim and Rees, 1992b).

Fig. 2.4. â–â interface of the MoFe protein as observed from the X-ray crystallographic structure of the MoFe protein

above proteins) indicates that these homodimers may function in a similar manner to that of the heterotetrameric native MoFe protein. These findings imply an *in vivo* communication between the NifD-K fusion proteins, which may be responsible for any of the processes involved in the normal functioning of the nitrogenase enzyme.

Acknowledgments We thank the members of Gavini/Pulakat laboratories at MSU for their helpful discussions and for critical reading of the manuscript. This research was supported in part by the National Institute of Health grant GM57636 and NSF grant to N. Gavini and L. Pulakat.

References

Blanchard, C. Z. and Hales, B. J. 1996. Isolation of two forms of the nitrogenase VFe protein from *Azotobacter vinelandii*. Biochemistry, 35(2): 472-478.

Brigle, K. E., Newton, W. E. and Dean, D. R. 1985. Complete nucleotide-sequence of the *Azotobacter vinelandii* nitrogenase structural gene-cluster. Gene, 37(1-3): 37-44.

Burgess, B. K. and Lowe, D. J. 1996. Mechanism of molybdenum nitrogenase. Chem. Rev., 96(7): 2983-3011.

Christiansen, J., Dean, D. R. and Seefeldt, L. C. 2001. Mechanistic features of the Mo-containing nitrogenase. Ann. Rev. Plant. Physiol. Plant. Mol. Biol.., 52: 269-295.

Dove, S. L. and Hochschild, A. 1998. Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. Genes & Dev., 12(5): 745-754.

Dove, S. L., Joung, J. K. and Hochschild, A. 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. Nature, 386(6625): 627-630.

Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J. and Rees, D. C. 1992. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. Science, 257(5077): 1653-1659.

Hofmann-Findeklee, C., Gadkari, D. and Meyer, O. 2000. Superoxide-dependent nitrogen fixation. In: Nitrogen Fixation- From Molecules to Crop Productivity. Kluwer, Boston, pp.23-30.

Howard, J. B. and Rees, D. C. 1996. Structural basis of biological nitrogen fixation. Chem. Rev., 96(7): 2965-2982.

Jang, S. B., Seefeldt, L. C. and Peters, J. W. 2000. Insights into nucleotide signal transduction in nitrogenase: Protein with MgADP bound. Biochemistry, 39(48): 14745-14752.

Kim, J. and Rees, D. C. 1994. Nitrogenase and biological nitrogen-fixation. Biochemistry, 33(2): 389-397.

Kim, J., Woo, D. and Rees, D. C. 1993. X-Ray crystal-structure of the nitrogenase molybdenum iron protein from *Clostridium pasteurianum* at 3.0 Angstrom Resolution. Biochemistry, 32(28): 7104-7115.

Kim, J. S., Rees, D. C. 1992a. Crystallographic structure and functional implications of the nitrogenase molybdenum iron protein from *Azotobacter vinelandii*. Nature, 360(6404): 553-560.

Kim, J. S., Rees, D. C. 1992b. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. Science, 257(5077): 1677-1682.

Loveless, T. M., Saah, J. R. and Bishop, P. E. 1999. Isolation of nitrogen-fixing bacteria containing molybdenum-independent nitrogenases from natural environments. Appl. Environ. Microbiol. 65 (9): 4223-4226.

Loveless, T. M. and Bishop, P. E. 1999. Identification of genes unique to Mo-independent nitrogenase systems in diverse diazotrophs. Can. J. Microbiol. 45(4): 312-317.

Mayer, S. M., Gormal, C. A., Smith, B. E. and Lawson, D. M. 2002. Crystallographic analysis of the MoFe protein of nitrogenase from a *nifV* mutant of *Klebsiella pneumoniae* identifies citrate as a ligand to the molybdenum of iron molybdenum cofactor (FeMoco). J. Biol. Chem., 277(38): 35263-35266.

Moshiri, F., Crouse, B. R., Johnson, M. K. and Maier, R. J. 1995. The nitrogenase-protective FesII protein of *Azotobacter vinelandii* - overexpression, characterization, and crystallization. Biochemistry, 34(40): 12973-12982.

Patil, P., Pulakat, L. and Gavini, N. Orf9 Functions as Nitrogenase-Specific ClpX During the Biogenesis of the MoFe-protein. 103rd General Meeting of American Society for Microbiology, Washington D.C., May 18-22, 2003, pp.83

Peters, J. W., Fisher, K. and Dean, D. R. 1995. Nitrogenase structure and function - a biochemical-genetic perspective. Ann. Rev. Microbiol., 49: 335-366.

Peters, J. W., Stowell, M. H. B., Soltis, S. M., Finnegan, M. G., Johnson, M. K. and Rees, D. C., 1997. Redox-dependent structural changes in the nitrogenase P-cluster. Biochemistry, 36(6): 1181-1187.

Rajagopalan, K.V. and Johnson, J. L., 1992. The pterin molybdenum cofactors. J. Biol. Chem, 267(15): 10199-10202.

Rees, D. C., Chan, M. K. and Kim, J. 1993. Structure and function of nitrogenase. Adv. Inorg. Chem., 40: 89-119.

Ribbe, M., Gadkari, D. and Meyer, O. 1997. N_2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-

superoxide oxidoreductase that couple N_2 reduction to the oxidation of superoxide produced from O_2 by a molybdenum-CO dehydrogenase. J. Biol. Chem., 272(42): 26627-26633.

Sambrook, J. F., Fritsch, E. F., and Maniatis,T. 1992. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Schindelin, N., Kisker, C., Sehlessman, J. L., Howard, J. B. and Rees, D. C. 1997. Structure of ADP x AIF(4)(-)-stabilized nitrogenase complex and its implications for signal transduction. Nature, 387(6631): 370-376.

Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B. and Rees, D. C. 1998. Conformational variability in structures of the nitrogenase iron proteins from *Azotobacter vinelandii* and *Clostridium pasteurianum.* J. Mol. Biol., 280(4): 669-685.

Stiefel, E. I. 1996. Molybdenum bolsters the bioinorganic brigade. Science, 272: 1599-1600.

Strange, R. W., Eady, R. R., Lawson, D. and Hasnain, S. S. 2003. XAFS studies of nitrogenase: the MoFe and VFe proteins and the use of crystallographic coordinates in three-dimensional EXAFS data analysis. J. Synchrotron. Radiat., 10: 71-75.

Stratagene, Inc. 2001. User manual for BacterioMatch^{TM} two-hybrid system. Stratagene, Inc., La Jolla, Calif.

Suh, M.H., Pulakat, L. and Gavini, N. 2003. Functional expression of a fusion-dimeric MoFe protein of nitrogenase in *Azotobacter vinelandii*. J. Biol. Chem., 278(7): 5353-5360.

CHAPTER III

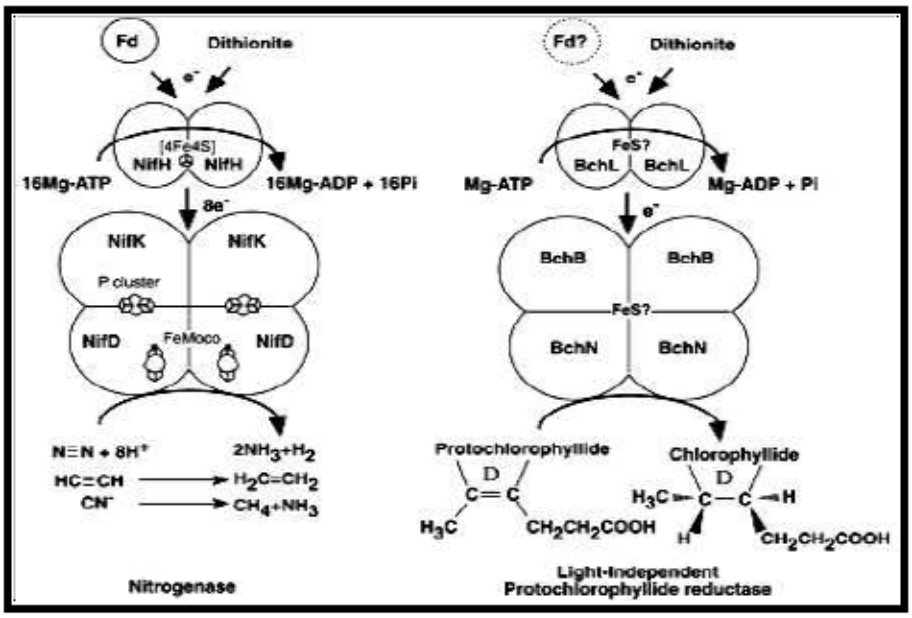
STRUCTURAL MODELING OF THE CHLN SUBUNIT OF THE LIGHT-INDEPENDENT PROTOCHLOROPHYLLIDE REDUCTASE BASED ON SIMILARITY WITH THE NIFD SUBUNIT OF THE NITROGENASE ENZYME

Introduction

The multi-subunit nitrogenase enzyme catalyzes the reduction of dinitrogen to ammonia. It consists of two separable components, the 240 kDa $\alpha_2\beta_2$ heterotetrameric MoFe (NifDK) protein encoded by the *nifD* and *nifK* genes and the 64 kDa homodimeric Fe (NifH) protein encoded by the *nifH* gene (for reviews see: Christiansen et al., 2001; Peters and Szilagyi, 2006). The Fe protein that contains a [4Fe-4S] metallic center is the obligate electron donor for the MoFe protein in the process of nitrogen fixation (Georgiadis et al., 1992). The MoFe protein contains the actual substrate reduction site in form of the metallic cluster known as the FeMo-cofactor (FeMoco) and therefore directly catalyzes the reduction of dinitrogen to ammonia (Kim and Rees, 1992a). Apart from the FeMoco, another metallic cluster known as the P-cluster is also present in the MoFe protein and is believed to be involved in the transfer of electrons from the [4Fe-4S] center to the FeMoco (Kim and Rees, 1992a; Kim and Rees, 1992b). The components of the light-independent protochlorophyllide reductase (DPOR) in cyanobacteria or plants, known as ChlL, ChlN and ChlB, share significant sequence similarity with the NifH,

NifD and NifK proteins of the nitrogenase complex respectively. At the primary sequence level, the ChlL and NifH share ~35% identity, the ChlB and NifK share ~19% identity and ChlN is homologous to both NifD and NifK by 19% (Fujita et al., 1993). The counterparts of the ChlL, ChlB and ChlN found in bacteria are known as BchL, BchB and BchN respectively. The sequence comparison between the NifH and ChlL clearly indicates the conservation of the ATP-binding motif (GXXXXGKS) and the cysteine ligands of the [4Fe-4S] cluster in ChlL. Fujita and Bauer proposed a model of the DPOR based on its similarity to the nitrogenase enzyme (Fujita and Bauer, 2000). They noted the obvious similarities between the NifH and ChlL protein and also indicated that since the BchN and BchB proteins could be co-purified in an equimolar ratio, there was strong possibility that the NB proteins formed a $(BchN)_2(BchB)_2$ heterotetramer similar to the $(NifD)_2(NifK)_2$ MoFe protein. Thus, similarities between the DPOR and nitrogenase subunits suggested that these complexes shared common molecular-structural architecture (Fig. 3.1). It is however true that the issue of the presence of putative P-cluster and FeMoco metallocenters that may be present in the ChlN-ChlB proteins still remains unsolved. The presence of Fe-S centers in the ChlB and ChlN proteins has not yet been resolved. This is because sufficient quantities of the purified DPOR enzyme have not been easily isolatable. However, it is likely that Fe-S centers do exist because, for example, ChlB and ChlN purified protein fractions exhibit a faint brown color that is constant with known spectral properties of proteins that contain an Fe-S cluster (Fujita and Bauer, 2000). It has also been observed that the DPOR activity is very sensitive to inhibition by oxygen (Fujita and Bauer, 2000), indicating a

characteristic similar to nitrogenase, in which the Fe-S center is the reason for oxygen sensitivity.



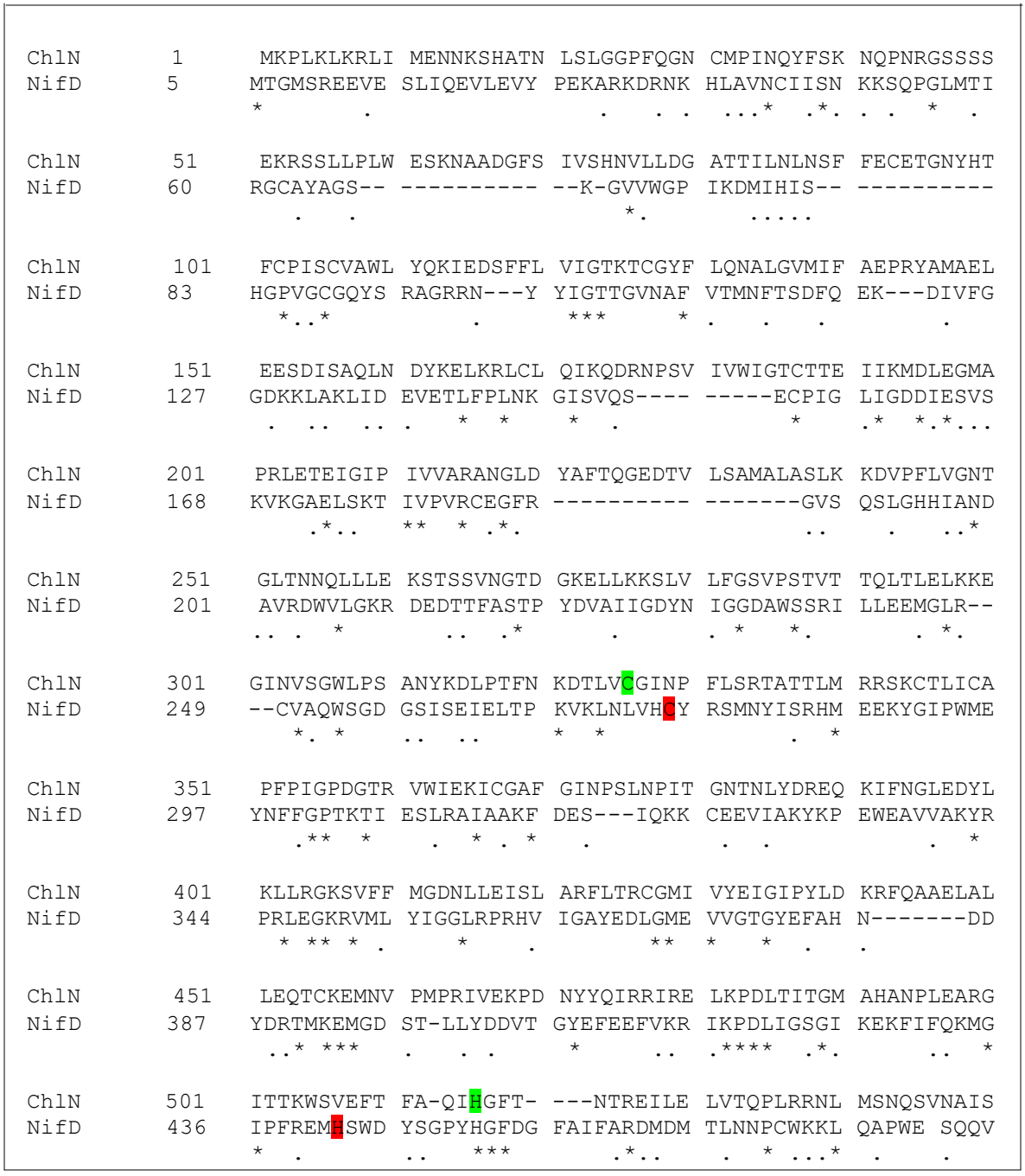
This model was based on the sequence similarity of DPOR to the Nitrogenase system.

Fig. 3.1. Similarities between the nitrogenase and DPOR components (Fujita and Bauer, 2000)

It is also possible that DPOR is actually more closely related to the NifE and NifN proteins of the nitrogenase system that are similar in constitution to the NifDK proteins and serve as a scaffold for the FeMoco precursor formation (Dos Santos et al., 2004). Studies have shown that NifEN contains two 4Fe-4S clusters instead of the two 8Fe:7S P clusters that are present in NifDK (Goodwin et al., 1998). Analysis of the Cys residues that serve as ligands for the P-cluster in the NifDK proteins showed that four of these were conserved in the ChlN and ChlB proteins (□-Cys88, □-Cys154, □-Cys70, □-Cys95),

implying that the N and B proteins might be involved in the formation of a 4Fe-4S redox center. The NifEN complex has a similar partial conservation of the P-cluster Cys residues (three in NifE and only one in NifN) (Brigle et al., 1987, Aguilar et al., 1990). As determined by primary sequence comparison data, no conservation of the residues that are involved in formation of the FeMo cofactor in nitrogenase was found in the ChlB / ChlN proteins, leading to the belief so far, that the protochlorophyllide (PChlide) reduction site is highly diverged from the dinitrogen reduction site in nitrogenase (Fujita et al., 1993; Fujita, 1996).

We were interested in applying structural homology modeling tools to determine the possibility of the presence of FeMoco in the ChlB or ChlN proteins. Mainly we analyzed the ChlB and ChlN sequences for optimum structural alignment with the NifD protein that contains the FeMoco, by using the 3-D X-ray crystallographic structure of the NifD protein as a structural template. The ChlB and ChlN protein sequences from *Chlamydomonas reinhardtii* were individually used for derivation of their 3-D structures based on their superimposition on the NifD protein from *Azotobacter vinelandii* (PDB ID: 3MIN (Peters et al., 1997)). As a part of our analysis, we also included the NifE and NifN protein sequences from *Azotobacter vinelandii* for the purpose of comparison of the residues that surround the FeMoco. We found better possibilities for the presence of FeMoco binding sites in the ChlN protein compared to the ChlB. The ChlN and NifD protein structures were structurally aligned or homology modeled based on the ClustalW sequence similarity of the two proteins (Fig. 3.2a). After the homology modeling of



The FeMoco ligands of NifD (*Azotobacter vinelandii*)are highlighted in red and the possible FeMoco ligands of ChlN (*Chlamydomonas reinhardtii*) are highlighted in green

Fig. 3.2a. ClustalW protein sequence alignment of NifD and ChlN.

ChlN based on NifD, we were surprised to note the presence of two possible counterparts that could serve as ligands for the FeMoco in ChlN. We were able to highlight the Cys326 and His515 residues of ChlN as ligands for the FeMoco since these residues closely coincided with the Cys275 and His442 ligands of the FeMoco found in NifD.

```
NifD    MTGMSREEVESLIQEVLEVYPEKARKDRNKHLAVNDPAVTQSKKCIISNKKSQPGLMTIR 60
NifE     -----------------------MKAKDIAELLDEPACSHNKKEKSGCAKPKPGATDGR 36
                                  * ::    :::** ::.**   .  *.:**    *

NifD     GCAYAGSKGVVWGPIKDMIHISHGPVGCGQYSRAGRRNYYIGTTGVNAFVTMNFTSDFQE 120
NifE     -CSFDG-AQIALLPVADVAHIVHGPIACAGSSWDNRGTRSSGPD----LYRIGMTTDLTE 90
          *:: *   :.  *: *: ** ***:.*.  *  .* .   *.     :  :.:*:*: *

NifD     KDIVFGGDKK--LAKLIDEVETLFPLNKGISVQSECPIGLIGDDIESVSKVKGAELSKTI 178
NifE     NDVIMGRAEKRLFHAIRQAVESYLPP--AVFVYNTCVPALIGDDVDAVCKAAAERFGTPV 148
         :*:::*  :*  :  : : **: :*   .: * . *  .*****:::*.*. . .:...:

NifD    VPVRCEGFRGVSQSLGHHIANDAVRDWVLGKRDEDTTFASTP------YDVAIIGDYNIG 232
NifE     IPVDSAGFYGTKN-LGNRIAGEAMLKYVIGTREPDPLPVGSERPGIRVHDVNLIGEYNIA 207
         :** . ** *..: **::**.:*: .:*:*.*: *.  ..:       :** :**:***.

NifD    GDAWSSRILLEEMGLRCVAQWSGDGSISEIELTPKVKLNLVHCYRSMNYISRHMEEKYGI 292
NifE    GEFWHVLPLLDELGLRVLCTLAGDARYREVQTMHRAEVNMMVCSKAMLNVARKLQETYGT 267
         *: *    **:*:*** :.  :**.   *::   :.::*:: * ::*  ::*:::*.**

NifD     PWMEYNFFGPTKTIESLRAIAAKFDES-IQKKCEEVIAKYKPEWEAVVAKYRPRLEGKRV 351
NifE    PWFEGSFYGITDTSQALRDFARLLDDPDLTARTEALIAREEAKVRAALEPWRARLEGKRV 327
         **:* .*:* *.* ::** :*  :*:. :  : * :**: :.: .*.:  :*.*******

NifD    MLYIGGLRPRHVIGAYEDLGMEVVGTGYEFAHNDDYDRTMKEMGDSTLLYDDVTGYEFEE 411
NifE    LLYTGGVKSWSVVSPLQDLGMKVVATGTKKSTEEDKARIRELMGDDVKMLDEGNARVLLK 387
         :** **::.  *:.. :****:**.** : : ::*  *  : ***.. : *: ..  : :

NifD     FVKRIKPDLIGSGIKEKFIFQKMGIPFREMHSWDYSGPYHGFDGFAIFARDMDMTLNNPC 471
NifE    TVDEYQADILIAGGRNMYTALKGRVPFLDINQEREFG-YGGYDRMLELVRHVCITLECPV 446
          *.. :.*:: :* :: :   *  :** :::.    * * *:* :  :.*.: :**: *

NifD     WKKLQ--APWEASEGAEKVAASA----- 492
NifE     WEAVRRPAPWDIPASQDARPSGGPFGER 474
         *: ::  ***: . . :  .:..
```

The FeMoco ligands of NifD (*Azotobacter vinelandii*)are highlighted in red and the possible FeMoco ligand of NifE (*Azotobacter vinelandii*) is highlighted in green

Fig. 3.2b. ClustalW protein sequence alignment of NifD and NifE

Also upon comparison of the important residues found in the surrounding region of the FeMoco in NifD (Dos Santos et al., 2004) with the superimposed region coinciding to this area in ChlN, close counterparts of several residues were found. We also detected a 'pocket-like' depression in the homology-modeled ChlN structure, similar to that found for the cavity that contains the FeMoco in NifD. We also performed protein sequence alignment for NifD and NifE (Fig. 3.2b) and performed homology modeling studies for NifE based on the NifD structural template. We found that in NifE, although a Cys residue corresponding to Cys275 of NifD could be observed, none of its His residues coincided with the His442 ligand of NifD. Thus, although additional structural and functional tools may be required to prove the actual presence of a FeMoco in ChlN, our structural based studies indicate a strong possibility for the same.

Materials And Methods

Multiple Alignments—Multiple alignments for the amino acid sequences of ChlB, ChlN, NifE, NifN and NifD were analyzed by ClustalW (Thompson et al., 1994). The ChlB and ChlN protein sequences (Accession numbers NP_958360 and NP_958412 respectively) were from *Chlamydomonas reinhardtii* and the NifE, NifN and NifD sequences (Accession numbers P08293, P10336 and P07328 respectively) were from *Azotobacter vinelandii*.

Secondary Structure Prediction—The secondary structure predictions of the ChlB, ChlN, NifE and NifN were performed using the Secondary Structure Prediction Tool (GOR4) of the San Diego Supercomputer Center (SDSC) (Garnier et al., 1978). This program

predicts four states, in which each residue is unambiguously assigned one conformational state of □-helix, extended chain, reverse turn or coil. Thus in a sample of 26 proteins, when the overall helix and extended-chain content is not taken into account, 49% of residue states are correctly predicted.

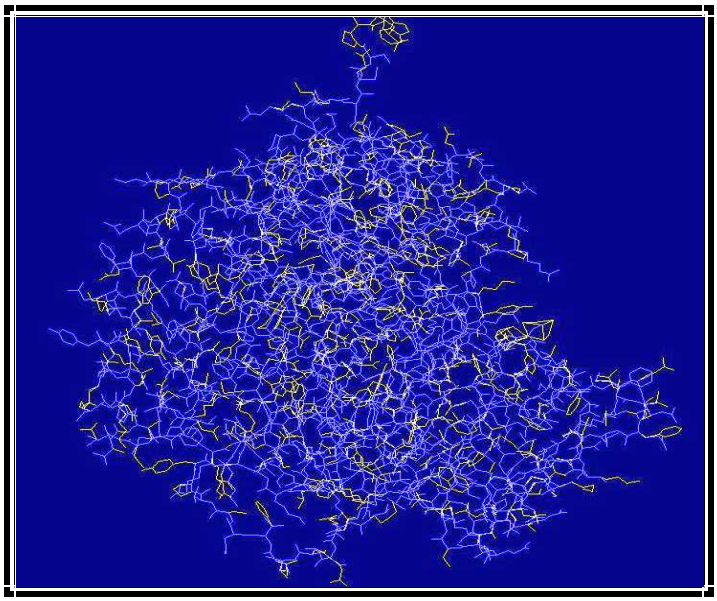
Molecular Modeling and Refined Graphical Representation—Homology models of ChlN and NifE were generated based on the available structural model of the NifD protein (Peters et al., 1997). The Swiss-PDB software (Guex and Peitsch, 1997) was used for performing the homology modeling in this study. We selected a strategy of semi-automatic and manual adjustments of the derived superimposed proteins relative to the available backbone conformation of the NifD protein structure. The structural alignment of the superimposed protein (for example, ChlB, ChlN, NifE, NifN) was done on basis of its CLUSTAL W sequence alignment with NifD. Finally, the images obtained were represented in a graphically refined format using the POV-Ray Imaging software (www.povray.org).

Phylogenetic analysis: Amino acid sequences of NifD, NifK, NifE, NifN, ChlB and ChlN were retrieved from the GenBank database. BLAST probing of the protein databases was performed with the BLASTP program (Altschul et al., 1997). Following are the GenBank accession numbers of the respective amino acid sequences: (1) NifD -- ABF06456.1 (*Nostoc sp*. 1189P), NP_907558.1 (*Wolinella succinogenes*), AAO38344.1 (*Leptospirillum ferooxidans*), P07328 (*Azotobacter vinelandii*); (2) NifK -- AAK33116.1 (*Methanosarcina mazei*), AAG60730.1 (*Bradyrhizobium japonicum*), NP_949952.1 (*Rhodopseudomonas palustris*), NP_435697.1 (*Sinorhizobium meliloti*), P11347

(*Clostridium pasteurianum*), P07329 (*Azotobacter vinelandii*), P25314 (*Azospirillum brasilense*); (3) NifE -- NP_444115.1 (*Rhizobium sp.* NGR234), AAT37645.1 (*Clostridium pasteuranium*), CAA40213.1 (*Bradyrhizobium japonicum*), P08293 (*Azotobacter vinelandii*), P08737 (*Klebsiella pneumoniae*), CAE30057.1 (*Rhodopseudomonas palustris*); (4) NifN -- NP_444116.1 (*Rhizobium sp*. NGR234), AAG60732.1 (*Bradyrhizobium japonicum*), AAN62899.1 (*Sinorhizobium meliloti*), P10336 (*Azotobacter vinelandii*), P08738 (*Klebsiella pneumoniae*); (5) ChlB -- NP_904211 (*Physcomitrella patens*), AAT28195 (*Larix decidua*), ABF60154.1 (*Stigeoclonium helveticum*), AAT80735 (*Huperzia lucidula*), YP_636477 (*Zygnema circumcarinatum*), YP_635732 (*Chara vulgaris*), BAB75140 (*Nostoc sp.* PCC 7120), BAA57916 (*Chlorella vulgaris*), P36437 (*Chlamydomonas reinhardtii*), P95463 (*Leptolyngba boryanum*), P48099 (*Cyanophora paradoxa*); (6) ChlN -- BAA18746 (*Synechocystis sp.* PCC 6083), NP_958412 (*Chlamydomonas reinhardtii*), AAT52199 (*Larix deciduas*), NP_848119 (*Adiantum capillus-veneris*), YP_764423 (*Stigeoclonium helveticum*), AAT80754 (*Huperzia lucidula*), YP_635867 (*Oltmannsiellopsis viridis*), YP_636520 (*Zygnema circumcarinatum*), YP_635806 (*Chara vulgaris*), BAE92312 (*Porphyra yezoensis*), NP_777473 (*Anthoceros formosae*), NP_817276 (*Pinus koraiensis*), NP_043286 (*Cyanophora paradoxa*), NP_045882 (*Chlorella vulgaris*), Q7NI15 (*Gloeobacter violaceus*), Q8YM64 (*Anabaena sp*. PCC 7120). The MEGA 2.1 software was used for the construction of phylogenetic trees using the maximum parsimony method (Kumar et al., 2004).

Results And Discussion

Preliminary analyses of the homology models. Although the ChlB and ChlN both had sequence similarity with NifD, it was found that the superimposed ChlN protein structure better suited our purpose for the study of FeMoco ligands and its surrounding environment. The ChlN model was built based on the structure derived by superimposition of the raw protein sequence onto the NifD structural template (PDB ID: 3MIN) (Fig. 3.3). The Swiss-PDB software allows us to better align two protein structures by manual adjustment of the sequence alignment that leads to the structural alignment. Our adjustments were based on the ClustalW sequence alignment data obtained by comparing the NifD and ChlN sequences.



The NifD structure (backbone) is shown in blue and the ChlN structure (backbone) is shown in yellow.

Fig. 3.3. Structural superimposition of derived ChlN structure on the NifD template.

We found that there were maximum identical residues in the following regions of the ChlN and NifD proteins: 346-351 (LEGKRV) NifD and 403-408 (LRGKSV) ChlN; 390-394 (TMKEM) NifD and 454-458 (TCKEM) ChlN; 417-421 (IKPDL) NifD and 481-485 (LKPDL) ChlN. We also took into account the predicted secondary structure for the ChlN protein while building its homology model (Fig. 3.4).

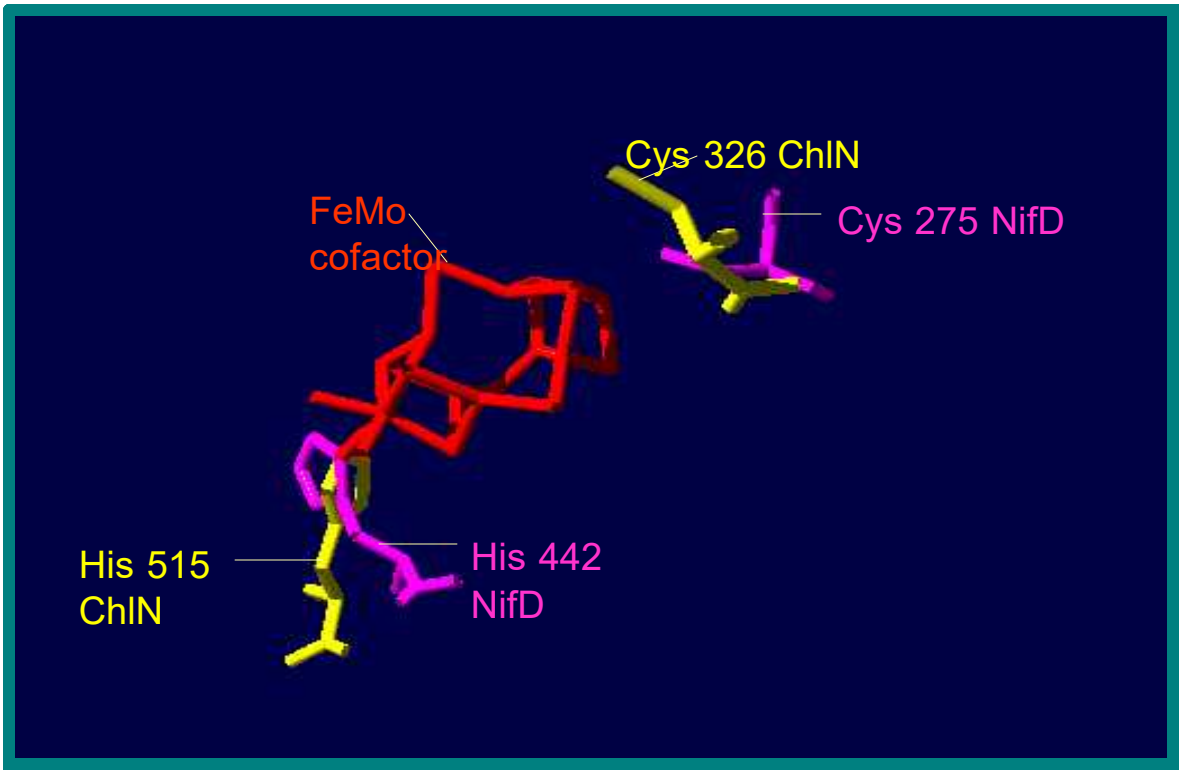
MKPLKLKRLIMENNKSHATNLSLGGPFQGNCMPINQYFSKNQPNRGSSSS
EKRSSLLPLWESKNAADGFSIVSHNVLLDGATTILNLNSFFECETGNYHT
FCPISCVAWLYQKIEDSFFLVIGTKTCGYFLQNALGVMIFAEPRYAMAEL
EESDISAQLNDYKELKRLCLQIKQDRNPSVIVWIGTCTTEIIKMDLEGMA
PRLETEIGIPIVVARANGLDYAFTQGEDTVLSAMALASLKKDVPFLVGNT
GLTNNQLLLEKSTSSVNGTDGKELLKKSLVLFGSVPSTVTTQLTLELKKE
GINVSGWLPSANYKDLPTFNKDTLVCGINPFLSRTATTLMRRSKCTLICA
PFPIGPDGTRVWIEKICGAFGINPSLNPITGNTNLYDREQKIFNGLEDYL
KLLRGKSVFFMGDNLLEISLARFLTRCGMIVYEIGIPYLDKRFQAAELAL
LEQTCKEMNVPMPRIVEKPDNYYQIRRIRELKPDLTITGMAHANPLEARG
ITTKWSVEFTFAQIHGFTNTREILELVTQPLRRNLMSNQSVNAIS

The alpha helices are shown in red, the beta sheets in blue and random coils in black.

Fig. 3.4. The secondary structure prediction of ChlN based on the GOR program of SDSC Workbench (Garnier et al., 1978).

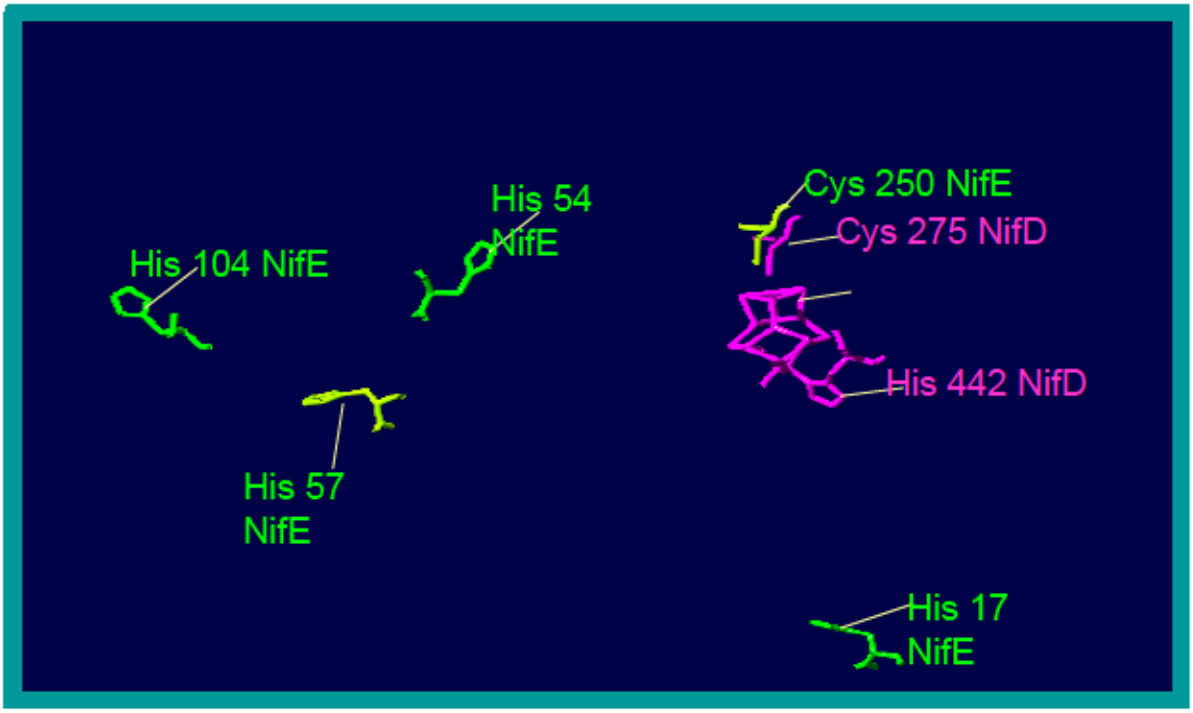
For the purpose of comparison of the derived ChlN structure with the NifE/NifN structures, we again used the NifD as a structural template and obtained the NifE and NifN homology models. It is known that the NifE sequence is more homologous to NifD than NifN is to NifD (Dean and Brigle, 1985; Fani et al., 2000). The NifE alongwith NifN is known to serve as a scaffold for the formation of the FeMoco preceding the FeMoco insertion stage into the MoFe protein. We therefore compared the NifD, ChlN and NifE structural models to find out the possibility of the presence of a FeMoco cluster in the ChlN protein, as found in NifD.

Analysis for the presence of FeMoco ligands in ChlN. The Cys275 and His422 residues have been shown to be the ligands for the FeMoco in NifD (Kim and Rees, 1992). As shown in Fig. 3.5a, it was observed that the Cys326 and His515 residues of ChlN were positioned similar to the Cys275 and His442 residues of the NifD, thus indicating the strong possibility for the presence of a FeMoco liganded by Cys326 and His515 in ChlN. In NifD the Cys275 is bound to the Fe at one end of the FeMoco cluster and the His442 is bound to the Mo at the opposite end of the cluster. The Mo is also coordinated by homocitrate (Kim and Rees, 1992). Recently, a FeMoco precursor was identified on the NifEN complex and it was suggested that this NifEN-bound precursor contained Fe as the only metal, and Mo and homocitrate were added while the cluster was still bound to the NifEN complex or at a later step (Hu et al., 2005). Sequence analyses of NifE also supported the same finding because although Cys250 of the NifE sequence corresponded to the Cys275 residue of NifD, a His residue corresponding to the His442 ligand of NifD was not present in NifE. Through structural homology modeling of NifE based on the NifD structural template, we examined the position of various His residues within the NifE protein that could serve as a counterpart to the His442 ligand of the NifD. However, we found that none of the His residues were positioned close to the presumed site of FeMoco in NifD (Fig. 3.5b). Therefore, in contrast to the findings in NifE, it was even more exciting to observe the possible structural conservation of both ligands for the FeMoco in ChlN.



The FeMoco cluster shown in red is present in the NifD protein and the Cys326 and His515 of ChlN are predicted to be the ligands for the presumed FeMoco based on superimposition of the derived ChlN structure on NifD.

Fig. 3.5a. Structural coincidence of FeMoco ligands (Cys275 and Cys 442) of NifD with Cys326 and His 515 of ChlN.



None of the His residues in NifE coincide with the His442 ligand of the FeMoco of NifD. The FeMoco cluster shown in red is present in the NifD protein and the Cys250 of NifE is predicted to be a ligand for the presumed FeMoco precursor based on superimposition of the derived NifE structure on NifD.

Fig. 3.5b. Cys 250 in NifE corresponds to the Cys275 ligand of NifD.

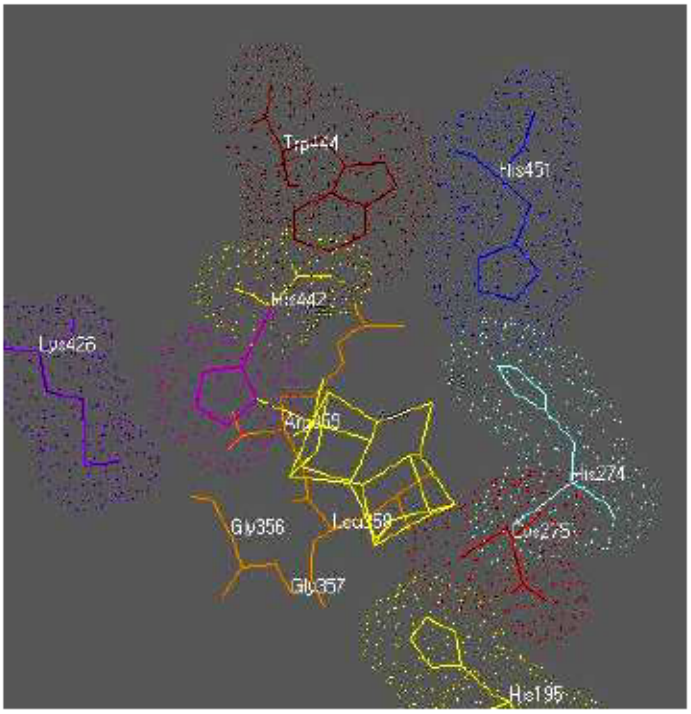
Important residues in the FeMoco environment in NifD and their counterparts in ChlN.

A comparison of the molecular surfaces of the apo-MoFe protein and the MoFe protein revealed the following important residues that may be involved in the FeMoco insertion (Dos Santos et al., 2004): (1) □-His^{274}, □ -His^{442}, □ -His^{451} ("His triad" in apo-MoFe protein); (2) □ -Trp^{444} (part of the FeMo-cofactor "lock"); (3) □ -355 through v -359 (part of the "lid" loop from □ -353 through □ -364); and (4) □ -Lys^{426} ("anchor" for the homocitrate of the FeMo-cofactor). We investigated the homology modeled ChlN

protein structure to detect if these important residues from NifD had possible counterparts in ChlN. It was observed that the following residues shared structural homology (Fig. 3.6a and 3.6b): Trp444 of NifD closely coincided with Trp362 of ChlN (Fig. 3.7); Residues 275-278 of ChlN showed structural coincidence with the residues 356-359 (part of the 'lid' loop) of NifD; Leu359 of NifD was situated similar to Leu275 in ChlN. However, no structurally homologous residues were found for the following residues of NifD in ChlN: His 451, Lys 426, His 274 and His 195. The His274 and His195 have been suggested to form a His triad with His442, helping to guide the negatively charged FeMoco during the insertion process (Dos Santos et al., 2004) but functional verifications have not followed yet. The □ -Trp444 residue has been recently shown to be involved in 'locking' the FeMoco in its binding site in the mature MoFe protein (Hu et al., 2006). The □ -Lys426 residue has been speculated to have a role in anchoring the FeMoco during the insertion event (Dos Santos et al., 2004).

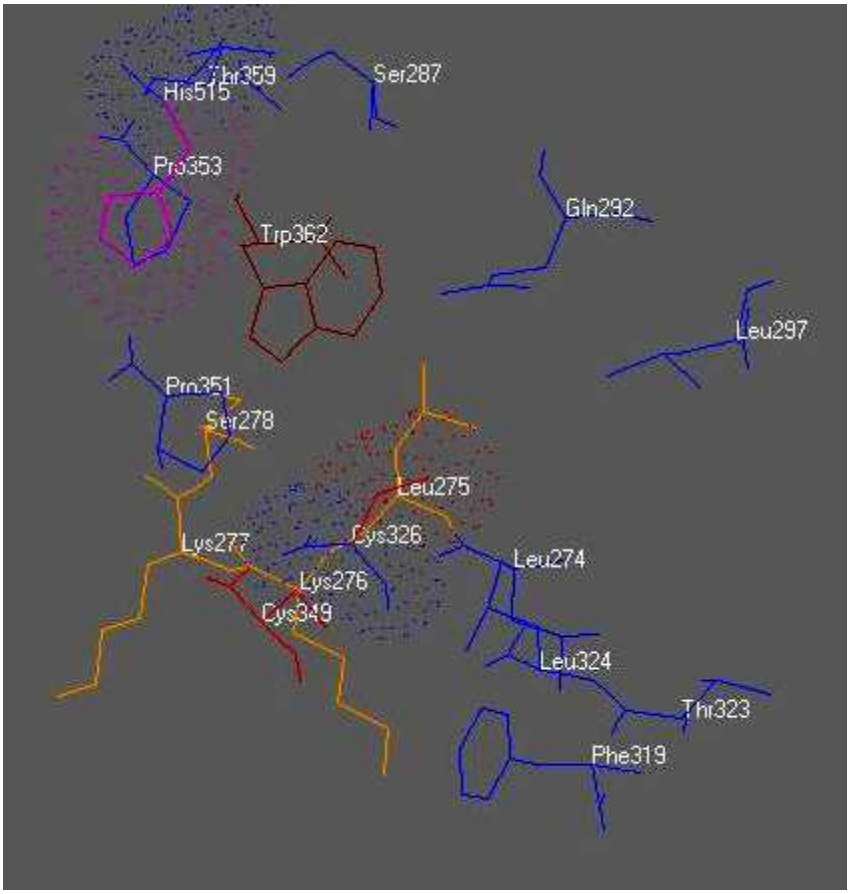
Presence of FeMoco pocket in ChlN? To further analyze the homology modeled ChlN structure in terms of the possibility of a FeMoco in it, we created a spacefilled model of NifD and ChlN and explored them for the presence of a FeMoco pocket. The NifD spacefilled model had a depression at the location of the FeMoco in its structure (Fig. 3.8a). It was interesting to note that the ChlN also had a pocket like appearance at the same region in its structure (Fig. 3.8b). This observation further supported the hypothesis that similar to NifD, a FeMoco cluster may be present in ChlN also. We however noted that the presumed FeMoco pocket in ChlN when compared to the corresponding pocket in NifD showed differences in terms of size and exposedness. The pocket in ChlN was

more wide open than that in NifD and it also appeared deeper than the NifD pocket (Fig. 3.8c and 3.8d).



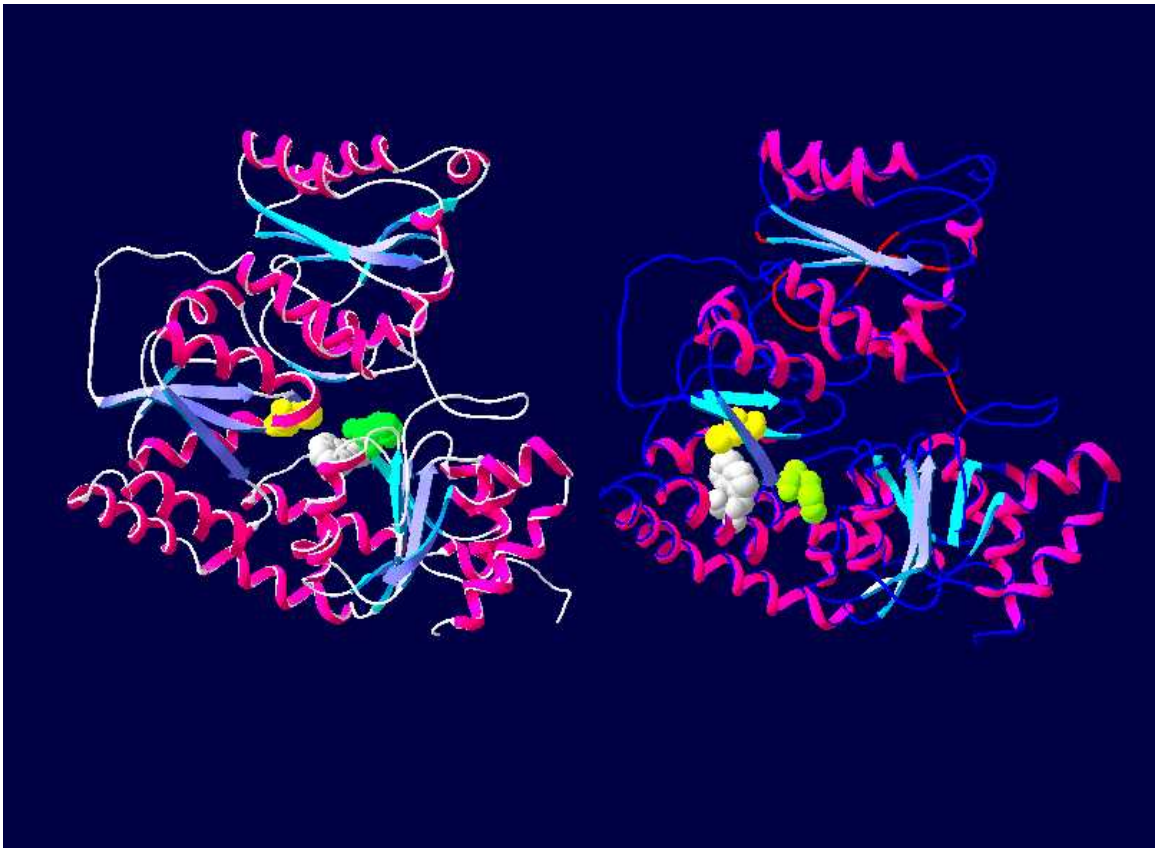
The key residues for FeMoco insertion include (1) □-His^{274}, □-His^{442}, □ -His^{451} ("His triad" in apo-MoFe protein); (2) □ -Trp^{444} (part of the FeMo-cofactor "lock"); (3) □ -355 through □ -359 (part of the "lid" loop from □-353 through □-364); and (4) □-Lys^{426} ("anchor" for the homocitrate of the FeMo-cofactor)

Fig. 3.6a. Important residues in the FeMoco environment in NifD (re-created from Dos Santos et al., 2004).



The residues in ChlN that structurally (based on superimposition) correspond to some of the key residues involved in FeMoco insertion in NifD are: (1) Trp362 (ChlN) □ Trp444 (NifD) (2) 275-278 (ChlN) □ 356-359 (NifD 'lid loop') (3) Cys326 (ChlN) □ Cys275 (NifD) (4) His515 (ChlN) □ His442 (NifD)

Fig. 3.6b. Important residues in the presumed FeMoco environment in ChlN.



The NifD model is shown on the left and the ChlN model is shown on the right. The alpha helices (ribbons) are in pink and the beta sheets are in light blue (ribbons). Residues Cys275 and Cys326 of NifD and ChlN respectively are shown in spacefill (yellow); Residues His442 and His515 of NifD and ChlN respectively are shown in spacefill (green); Residues Trp444 and Trp362 of NifD and ChlN respectively are shown in spacefill (gray). Red coils in the ChlN structure represent areas of poor structural alignment with NifD.

Fig. 3.7. Homology modeled ChlN structure depicting counterparts to Cys275, Trp444 and His442 of NifD.

Phylogenetic analysis of NifD, NifK, NifE, NifN, ChlB and ChlN: The evolutionary history of the nitrogenase and DPOR systems has remained a major subject of investigation. In case of the evolutionary relationships of NifD, NifK, NifE and NifN, studies have proposed that an ancestral gene gave rise to a bicistronic operon by an in-

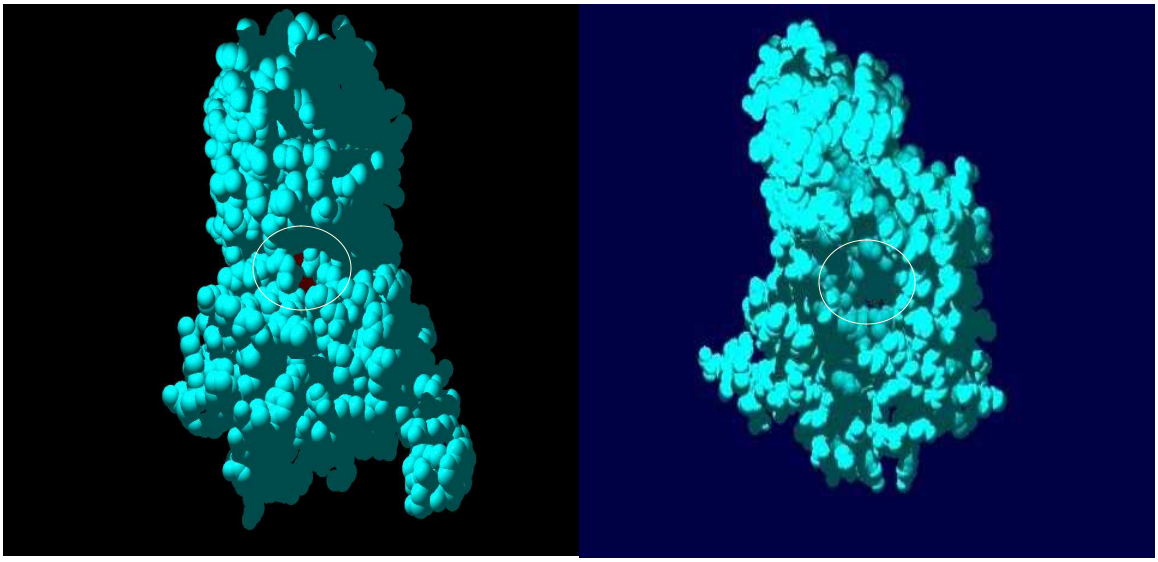


Fig. 3.8a (left) Spacefilled model of the NifD protein showing the site of FeMoco insertion.

Fig. 3.8b (right) Spacefilled model of the derived ChlN protein structure showing a pocket like appearance.

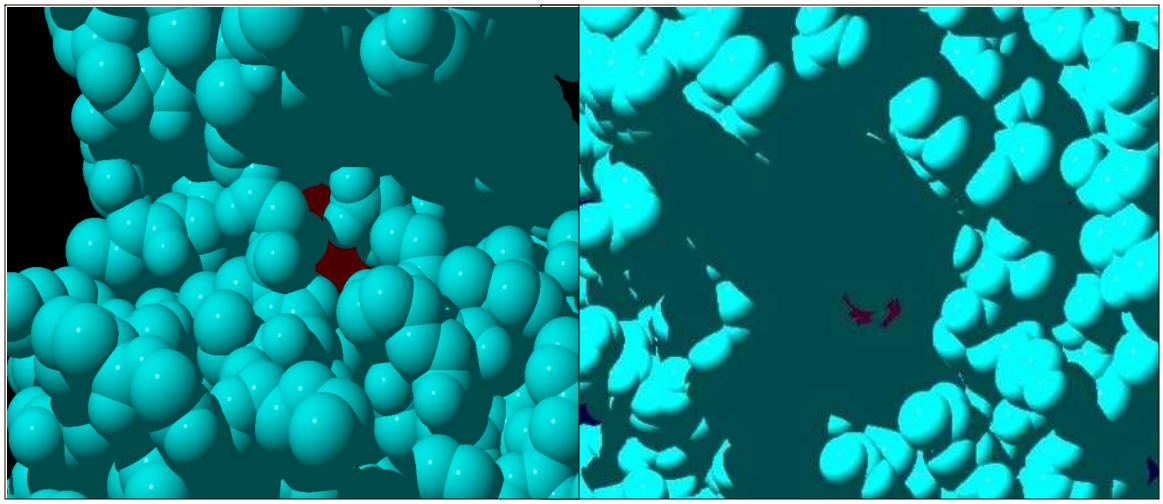
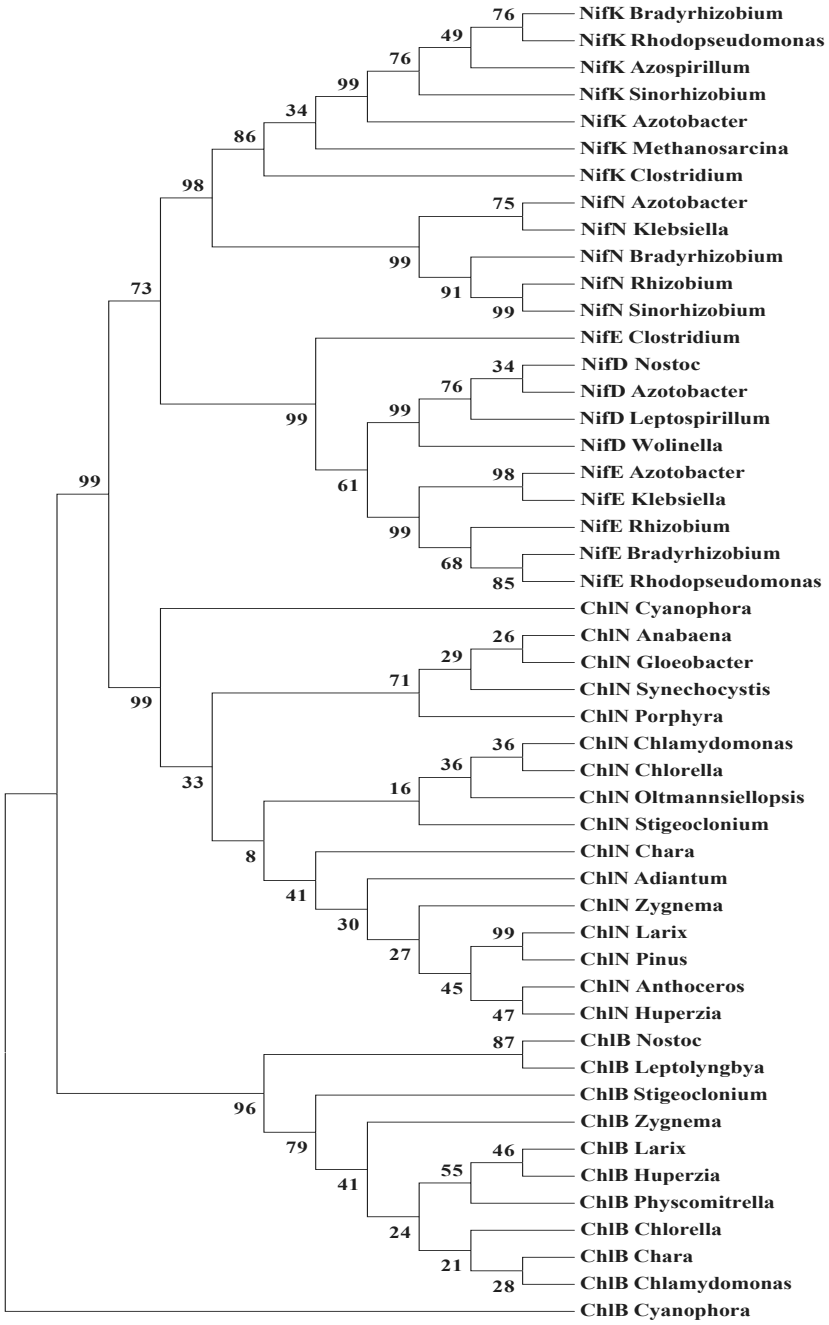


Fig. 3.8c (left) Magnified image of the spacefilled NifD model in Fig. 3.8a.

Fig. 3.8d (right) Magnified image of the spacefilled ChlN model in Fig. 3.8b.

tandem paralogous duplication event, followed by evolutionary divergence leading to ancestors of the present day *nifDK* and *nifEN* operons (Fani et al., 2000). We generated a phylogenetic tree using the NifD, NifK, NifE, NifN, ChlB and ChlN amino acid sequences. The phylogenetic tree was constructed based on the maximum parsimony method, available in the MEGA 3.1 (Kumar et al., 2004) phylogenetic software. As seen in Fig. 3.9, the phylogenetic tree pattern indicated that the ChlN protein was more closely related to the NifD, NifK, NifE and NifN proteins than ChlB. The bootstrap values were also calculated for each branch of the tree and most major nodes were found to have values >95%. As suggested by previous reports (Fani et al., 2000), the NifD and NifE proteins formed one sub-group while the NifK and NifN proteins formed another sub-group within the nitrogenase components in the phylogenetic tree created in our study. According to this phylogenetic tree, the ChlB protein may have been the precursor molecule for the subsequent evolution of the ChlN protein and the NifDKEN proteins. The aim of this study was to explore the homology modeled structure of the ChlN protein and examine if there were potential FeMoco binding or surrounding residues, as compared to those found in the NifD protein of the nitrogenase complex. Structural homology modeling based studies such as this have recently been performed by other groups wherein the Lhca4 subunit of LHCI-730 peripheral antenna in photosystem I was structurally modeled based on similarity with LHCII (Melkozernov and Blankenship, 2003) and the *E. coli* 4-hydroxybenzoic acid oligoprenyltransferase (ubiA transferase) was modeled for potential active sites (Brauer et al., 2004).



This tree was generated using the maximum parsimony method and bootstrap consensus values (based on 100 trees) calculated for each node of the tree (as indicated by the numbers).

Fig. 3.9. Phylogenetic tree highlighting ancestral relationships of the NifD, NifK, NifE, NifN, ChlB and ChlN.

Our structural modeling highlighted the Cys326 and His515 residues of the ChlN proteins as potential FeMoco ligands, by homology modeling studies based on the NifD structural template. Based on homology modeling, we have also indicated the possibility of a few other FeMoco surrounding residues in NifD to be present in ChlN.

References

Aguilar, O. M., Taormino, J., Thony, B., Ramseier, T., Hennecke, H. and Szalay, A. A. 1990. The *nifEN* genes participating in FeMo cofactor biosynthesis and genes encoding dinitrogenase are part of the same operon in *Bradyrhizobium* species. Mol Gen. Genet. 224(3): 413-420

Altschul, S., F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402

Brauer, L., Brandt, W. and Wessjohann, L. A. 2004. Modeling the *E. coli* 4-hydroxybenzoic acid oligoprenyltransferase (ubiA transferase) and characterization of potential active sites. J. Mol. Model (Online) 10(5-6): 317-327

Brigle, K. E., Weiss, M. C., Newton, W. E. and Dean, D. R. 1987. Products of the iron-molybdenum cofactor-specific biosynthetic genes, *nifE* and *nifN* are structurally homologous to the products of the nitrogenase molybdenum-iron protein genes, *nifD* and *nifK*. J. Bacteriol. 169(4): 1547-1553

Christiansen, J., Dean, D.R. and Seefeldt, L.C., 2001. Mechanistic features of the Mo-containing nitrogenase. Ann. Rev. Plant. Physiol. Plant. Mol. Biol. 52: 269-295.

Dean, D. R. and Brigle, K. E. 1985. *Azotobacter vinelandii nifD*- and *nifE*- encoded polypeptides share structural homology. Proc. Natl. Acad. Sci. USA. 82(17): 5720-5723

Dos Santos, P.C., Dean, D. R., Hu, Y. and Ribbe, M. W. 2004. Formation and Insertion of the Nitrogenase Iron-Molybdenum Cofactor. Chem. Rev. 2004; 104(2): 1159 – 1174

Fani, R., Gallo, R. and Pietro, L. 2000. Molecular evolution of nitrogen fixation: The evolutionary history of the *nifD*, *nifK*, *nifE* and *nifN* genes. J. Mol. Evol. 51: 1-11

Fujita, Y. 1996. Protochlorophyllide reduction: a key step in the greening of plants. Plant Cell. Physiol. 37(4): 411-421

Fujita, Y. and Bauer, C. E. 2000. Reconstitution of light-independent protochlorophyllide reductase from purified BChL and BchN-BchB subunits. *In vitro* confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. J. Biol. Chem. 275(31): 23583-23588

Fujita, Y., Matsumoto, H., Takahashi, Y. and Matsubara, H. 1993. Identification of a *nifDK*-like gene (ORF467) involved in the biosynthesis of chlorophyll in the cyanobacterium *Plectonema boryanum*. Plant Cell. Physiol. 34: 305-314

Garnier, J. Osguthorpe, D.J. and Robson, B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97-120

Georgiadis, M.M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J.J. and Rees, D.C., 1992. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii.* Science. 257(5077): 1653-1659.

Goodwin, P. J., Agar, J. N., Roll, J. T., Roberts, G. P., Johnson, M. K. and Dean, D. R. 1998. The *Azotobacter vinelandii* NifEN complex contains two identical [4Fe-4S] clusters. Biochemistry. 37(29): 10420-10428

Guex, N. and Peitsch, M. C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 18(15): 2714-2723

Hu, Y., Fay, A. W. and Ribbe, M. W. 2005. Identification of a nitrogenase FeMo cofactor precursor on NifEN complex. Proc. Natl. Acad. Sci. USA. 102(9): 3236-3241

Hu, Y., Fay, A. W., Schmid, B., Makar, B., Ribbe, M. W. 2006. Molecular insights into nitrogenase FeMoco insertion- Trp 444 of MoFe protein alpha subunit locks FeMoco in its binding site. J. Biol. Chem. [Epub ahead of print]

Kim, J.S., Rees, D.C., 1992a. Crystallographic structure and functional implications of the nitrogenase molybdenum iron protein from *Azotobacter vinelandii*. Nature 360(6404): 553-560.

Kim, J.S., Rees, D.C., 1992b. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. Science 257(5077): 1677-1682.

Kumar, S., Tamura, K. and Nei, M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163.

Melkozernov, A. N. and Blankenship, R. E. 2003. Structural modeling of the Lhca4 subunit of LhcI-730 peripheral antenna in photosystem I based on similarity with LhcII. J. Biol. Chem. 278(45): 44542-44551

Peters, J. W., Stowell, M. H., Soltis, S. M., Finnegan, M. G., Johnson, M. K. and Rees, D. C. 1997. Redox-dependent structural changes in the nitrogenase P-cluster. Biochemistry. 36: 1181-1187

Peters, J. W. and Szilagyi, R. K. 2006. Exploring new frontiers of nitrogenase structure and mechanism. Curr. Opin. Chem. Biol. 10(2): 101-108

Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22(22): 4673-4680

CHAPTER IV

THE NIFX PROTEIN IS INVOLVED IN THE FINAL STAGES OF FEMO-COFACTOR TRANSPORT TO THE MOFE PROTEIN

Introduction

The nitrogenase enzyme converts the atmospheric nitrogen into a bioavailable form, thus serving as a focal point in structural, molecular and biochemical studies (Igarashi and Seefeldt, 2003(*review*); Rees et al., 2005 (*review*); Barney et al., 2006). The molybdenum containing nitrogenases (Mo-nitrogenases) are the best characterized metalloenzymes and are composed of two oxygen sensitive metalloproteins known as the MoFe protein (NifDK) and the Fe protein (NifH). The MoFe protein contains two types of metal centers, the P-cluster and the iron-molybdenum cofactor (FeMoco) (Howard and Rees, 1996 (*review*)). The FeMoco is the active site of the Mo-nitrogenase and composed of seven Fe, nine S, one Mo, one homocitrate, and one atom of unidentified nature (Chan et al., 1993; Einsle et al., 2002). It is entirely situated in the NifD subunit which provides one Cys and one His ligand for bonding with the Fe and Mo ends of its metallostructure respectively. Considerable research efforts have been directed towards understanding the biosynthesis and sequential assembly of the FeMoco and it is known that various nitrogen fixation (*nif*) genes play a role in the biosynthesis, assembly,

transport and insertion of the FeMo-cofactor (Dos Santos et al., 2004; Rubio and Ludden, 2005). The products of the *nifB*, *nifV*, *nifQ*, *nifH*, *nifE*, *nifN*, *nifX* and *nafY* genes have been identified to be the prime players in these processes (Peters et al., 1995(*review*); Shah et al., 1999; Rubio et al., 2004).

Two different pathways for the biosynthesis of FeMoco could be pictured: (i) the FeMoco units could be sequentially assembled on the apo-MoFe or (ii) the FeMoco could be separately synthesized and then inserted onto the apo-MoFe protein. However, as established by past experiments, the second pathway proved to be accurate, whereby mutations in either *nifB*, *nifN* or *nifE,* that are involved in the biosynthesis of FeMoco, resulted in very low molybdenum accumulation and a molybdenum-free MoFe protein (Imperial et al., 1987; Ugalde et al., 1984). Thus it was demonstrated that an active FeMoco could be synthesized in the absence of the MoFe protein and that the FeMoco could accumulate on a different protein, such as an intermediate in the normal FeMoco biosynthetic pathway (Ugalde et al., 1984). A significant progress in determining the protein components involved in the biosynthesis of the FeMoco ensued from the development of an *in vitro* assay for FeMoco biosynthesis (Shah et al., 1986). This assay utilized a mixture of extracts prepared from mutant strains that presumably had complementary defects in FeMoco biosynthesis, so as to attempt reconstitution of the MoFe protein activity. This application proved immensely useful in recognizing NifH, NifEN, and NifV as some of the key players in the FeMoco biosynthetic pathway and helped in the purification of an apo-MoFe protein (Ugalde et al., 1984, Shah et al., 1986; Robinson et al., 1987; Paustian et al., 1989; Meijer and Tabita, 1992). As understood

from research conducted thus far, a stepwise overview of the events involved in the biosynthesis and assembly of the FeMoco includes: (i) the formation of an Fe/S core of the FeMoco (designated NifB-co) by NifB (Shah et al., 1994; Allen et al., 1995) (ii) transfer of the NifB-co to the □$_2$□$_2$ tetrameric scaffold NifEN protein (Goodwin et al., 1998; Roll et al., 1995) (iii) maturation of the FeMoco and apo-MoFe protein by the combined action of the Fe protein and MgATP (Robinson et al., 1987; Allen et al., 1996; Chatterjee et al., 1994) and further processing of the NifB-co on NifEN by an unknown mechanism, forming the completed FeMoco (iv) the final transfer of the fully formed FeMoco to a *nafY*-encoded protein, called □which is a FeMoco carrier that aids in the insertion of the FeMoco into the apo-MoFe (Homer et al., 1995; Rubio et al., 2002). In addition to these major events, a few other important steps are also involved in the entire process. The NifU and NifS proteins participate in the initial Fe/S mobilization (Yuvaniyama et al., 2000; Zheng et al., 1997), the NifQ protein is involved in an early stage of the metal core formation and found to be essential in Mo-limiting conditions (Imperial et al., 1984), the NifV protein contributes the homocitrate component of the FeMoco (Kennedy and Dean, 1992) and the NifX protein probably functions as an 'escort' protein that delivers the FeMoco or its precursors from one assembly site to another or plays a role in specifying the organic acid moiety of FeMoco (Rangaraj et al., 2001; Rangaraj and Ludden, 2002.

Although a wealth of data is available to support the functions of the NifU, NifS, NifB, NifEN, NifH and NifV, more research is required to explain the role of the NifX protein. The *nifX* gene is located downstream of the *nifEN* genes, such that *nifENX* are

transcribed as a single transcription unit. The NifX protein is comprised of 158 amino acids and is a ~17 kDa protein (http://ca.expasy.or/uniprot/14887). The small molecule binding domain (SMBD) of NifX has widespread phyletic distribution and therefore indicates an ancient origin or extensive dissemination by horizontal transfer during evolution, or both (Anantharaman et al., 2001). Eventhough a deletion in the *nifX* gene did not show a significant effect on the nitrogenase activity *in vivo*, the purified NifX protein was shown to be important in the stimulation of FeMoco synthesis in an *in vitro* FeMoco synthesis reaction (Shah et al., 1999; Jacobson et al., 1989). The NifX, NifB, NifY and NafY are most likely related proteins, owing to their overall sequence similarity and the presence of a FeMoco binding 'core domain' in all of them (Dyer et al., 2003). This conserved family of proteins has a □$_1$-□$_2$-□$_3$-□$_1$-□$_2$-□$_4$□□$_3$-□$_5$-□$_4$-□$_5$ fold in its core domain, a polarized surface charge distribution, and shows considerable similarity with the ribonuclease H family (Dyer et al., 2003). In NifB, the NIFX SMBD is fused with a biotin-synthase-like, metal-dependent catalytic domain, and in the archaeal MTH1172 protein, it is fused to a cation transporter (Anantharaman et al., 2001). Previous studies in *Azotobacter vinelandii* have identified certain important molecular interactions of the NifX protein with the other Nif proteins involved in the FeMoco biosynthesis/transport pathway; mainly it was shown that NifX binds to the NifB-co and also a FeMoco precursor from the NifEN complex (Shah et al., 1999; Rangaraj et al., 2001).

Clearly, a protein-protein interaction based study of NifX with other FeMoco maturation proteins would enhance our insight into the function of NifX and may help to solve unanswered questions regarding its role in the FeMoco biosynthetic pathway. We

therefore studied the interactions of NifX with some of the other relevant proteins participating in FeMoco biosynthesis /assembly/transport /insertion, so as to deduce the entire sequence of events involved. We determined the specific interactions of NifX with NifB, NifH, NifN, NifD, NifK, NifDK (fusion), and NafY using the BacterioMatch Two-Hybrid System™ (Dove et al., 1997, BacterioMatch™ Manual, 2001). Accordingly, a translationally fused construct of NifX with the N-terminal □-RNAP of the pTRG target vector was made and its interaction was tested with the above proteins, each translationally fused to the ëCI of the pBT vector. The strength of the interactions, as determined by measuring the â-galactosidase activity, demonstrated that direct protein-protein interaction exists between (i) NifB and NifX (ii) NafY and NifX and (iii) NifDK (fusion) and NifX (iv) NifK and NifX. The extent of interaction between NifK and NifX proteins was much higher than between NifD and NifX, when individually tested; also, weak interaction was found between NifH and NifX and NifN and NifX. Our results present a new dimension to the FeMoco biosynthesis pathway with reference to the role played by NifX. We propose that the NifX performs overlapping functions in the FeMoco biosynthetic pathway wherein it may either function as a transporter of the FeMoco precursor to the NafY protein or directly assist in the insertion of the completed FeMoco into the apo-MoFe protein.

Materials And Methods

Strains, plasmids, and growth conditions. The bacterial strains and plasmids used in this study are described in Table 4.1. *E. coli* strains were normally grown at 37 °C in 2YT media (Sambrook et al., 1992). Ampicillin, chloramphenicol and tetracycline were used

Table 4.1. Bacterial strains and plasmids used in this study

Strain/plasmid	Relevant characteristics and description	Source/reference
Escherichia coli TOP10	F′*mcr*A □ (*mrr-hsd*RMS-*mcr*BC) □ 80*lacZ*□M15 □*lac*X74 *rec*A1 *ara*D139 □(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen, CA
Escherichia coli XL-1 Blue	*MRF′ K, Ä(mcrA) 183Ä(mcr CB-hsdSMR-mrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac (F′ pro AB lacI^q Z Ä M15 Tn5 (Kan)′]*	Stratagene, CA
PCR 2.1 TOPO	*Amp′, Kan′,* (3.9kb), used for direct cloning of PCR products, *lacZá* fragment, *MCS, M13*	Invitrogen, CA
pBT	*Cm^R*, 52 bp *MCS*, 3.2 kb size, *MCS, p15A* origin of replication, *lac-UV5* , ë cI ORF	Stratagene, CA
pTRG	*Tet^R*, 60 bp *MCS*, 4.4 kb size, MCS, *lac-UV5* promoter, *CoE1* origin of replication, *RNAPá ORF*	Stratagene, CA
pBG1866	Derivative of pBT in which 879 bp *nifH* fragment was cloned into *BamHI* site of pBT to generate an in-frame ëCI:NifH translation fusion	(Verma, 2002)
pBG1713	Derivative of pBT in which 3062 bp *orf* corresponding to *nifD-K* fusion gene from *XhoI-EcoRV* digested pBG1712 was cloned into *SmaI-XhoI* site of pBT bait vector to generate an in-frame ëCI:NifD-K translation fusion	(Lahiri et al., 2005)
pBG1716	Derivative of pTRG in which 2335 bp nifKTY fragment was cloned into the *BamHI* site of pTRG to generate an in-frame á-RNAP: NifX translation fusion	(Lahiri et al., 2005)
pBG1718	Derivative of pBT in which 2335 bp *nifKTY* fragment released from *BamHI* digested pBG1715 was cloned into *BamHI* site of pBT to generate an in-frame ëCI NifKTY translation fusion	(Lahiri et al., 2005)
pBG1777	Derivative of PCR 2.1 TOPO in which 1377 bp *nifN* fragment was cloned directly after PCR using appropriate oligonucleotides designed with flanking *EcoRI-BamHI* sites	This study
pBG1778	Derivative of pBT in which 1377 bp *nifN* fragment released from *EcoRI-Bam*HI digested pBG1777 was cloned into the *EcoRI-BamHI* site of pBT to generate an in-frame ëCI:NifN translation fusion	This study
pMH5014	Derivative of PCR 2.1 TOPO in which 1-300 bp *nifX* fragment was cloned directly after PCR using appropriate oligonucleotides designed with flanking *BamHI-EcoRI* sites	This study
pBG1786	Derivative of pTRG in which 1-300 bp *nifX* fragment released from *BamHI-EcoRI* digested pMH5014 was cloned into the *BamHI-EcoRI* site of pTRG to generate an in-frame □-RNAP:N1NifX translation fusion	This study
pBG1782	Derivative of PCR 2.1 TOPO in which 150-477 bp *nifX* fragment was cloned directly after PCR using appropriate oligonucleotides designed with flanking *BamHI-EcoRI* sites	This study
pBG1787	Derivative of pTRG in which 150-477 bp *nifX* fragment released from *BamHI-EcoRI* digested pBG1782 was cloned into the *BamHI-EcoRI* site of pTRG to generate an in-frame □-RNAP:C2NifX translation fusion	This study
pMH5000	Derivative of PCR 2.1 TOPO in which 477 bp *nifX* fragment was cloned directly after PCR using appropriate oligonucleotides designed with flanking *BamHI-EcoRI* sites	This study
pMH5002	Derivative of pTRG in which 477 bp *nifX* fragment released from *BamHI-EcoRI* digested pMH5000 was cloned into the *BamHI-EcoRI* site of pTRG to generate an in-frame □-RNAP:NifX translation fusion	This study
pMH6000	Derivative of PCR 2.1 TOPO in which 1509 bp *nifB* fragment was cloned directly after PCR using appropriate oligonucleotides designed with flanking *EcoRI-BamHI* sites	This study
pMH6002	Derivative of pBT in which 1509 bp *nifB* fragment released from *EcoRI-BamHI* digested pMH6000 was cloned into the *EcoRI-BamHI* site of pBT to generate an in-frame ëCI:NifB translation fusion	This study
pBG1785	Derivative of PCR 2.1 TOPO in which 732 bp *nafY* fragment was cloned directly after PCR using appropriate oligonucleotides designed with flanking *EcoRI-BamHI* sites	This study
pBG1790	Derivative of pBT in which 732 bp *nafY* fragment released from *EcoRI-BamHI* digested pBG1785 was cloned into the *EcoRI-BamHI* site of pBT to generate an in-frame ëCI:NafY translation fusion	This study

to a final concentration of 50, 34 and 5 μg/ml,respectively, wherever the selection was made.

General molecular techniques. Restriction enzymes were purchased from Promega (Madison, WI). DNA sub-cloning, plasmid DNA isolations, restriction enzyme digestions, agarose gel electrophoresis, ligations, and *E. coli* transformation were carried out as described in the laboratory manual or according to the manufacturer's protocol (BacterioMatch™ Manual, 2001; Sambrook et al., 1992) . Oligonucleotides used for PCR amplification were purchased from Integrated DNA Technologies (Coralville, IA). Following are the primers used to PCR amplify required gene fragments from *A. vinelandii* cells for cloning into pTRG or pBT vectors of the BacterioMatch Two Hybrid System™: (1) *nifX* (pTRG) (i) 5' GGA TCC ATG TCC AGC CCG ACC CGA CAA TTG CAG GTA 3' and (ii) GAA TTC CTA TTC GTC CCA GCC TTC GGC GGC CAT GGC 3'. (2) *nifB* (pBT) (i) 5' GAA TTC AGA ACT GAG CGT ACT TGG GCA AAA CAA T 3' and (ii) 5'GGA TCC TCA GGC CTT GGC CTG CAG CAG GGC T 3' (3) *nifN* (pBT) (i) 5'GAA TTC ATG GCC GAG ATC ATC AAT CGC AAC AAG GCC 3' and (ii) 5' GGA TCC TCA GTG CCT CCA TTG CGG CTG TTC GGT TGC CGG 3' (4) *nafY* (pBT) (i) 5'GAA TTC AGT AAC CCC CGT GAA CAT GAG TCG CGA 3' (ii) 5' GGA TCC TCA TGC CCT GGC CGC CTC GTC CTC GTC 3' (5) First 300 bp of *nifX* encoding N-terminal NifX (pTRG) (i) 5' GGA TCC ATG TCC AGC CCG ACC CGA CAA TTG CAG GTA 3' (ii) 5' GAA TCC GGC CAT CAA CTG GCG CAC CGC CGA GGC GCC 3' (6) Last 324 bp of *nifX* encoding C-terminal NifX (pTRG) (i) 5'GGA TCC CGC TCC CAG CTC CTC TCG GTC GTC GAG TTC 3' (ii) GAA TTC

CTA TTC GTC CCA GCC TTC GGC GGC CAT GGC 3'. For insertion of a gene in-frame with the 5' *EcoRI* site of the pBT plasmid, oligonucleotides were designed with an extra nucleotide at their start. In all the above oligonucleotides, the underlined sequences represent restriction enzyme recognition sites that were introduced during primer design so as to facilitate cloning of the DNA fragments into unique sites of the pBT or pTRG vectors. 'GGA TCC' is the *BamHI* recognition sequence and 'GAA TTC' is the *EcoRI* recognition sequence. The bacterial two-hybrid vectors pBT bait and pTRG target, containing the ëCI and á-RNAP domains, respectively, were the starting plasmids used for cloning the various gene fragments. Each gene fragment was PCR amplified using specific oligonucleotides as listed above, and cloned into PCR 2.1 TOPO (Invitrogen, Carlsbad, CA) by ligation and transformation. These plasmids were then digested with *EcoRI* and *BamHI* and the resulting fragments were purified and ligated to the *EcoRI-BamHI* digested pBT or pTRG vectors. The ligated mixtures were used to transform *E. coli* XL1-Blue cells and desired clones were obtained and verified from the subsequent transformants by DNA isolation and restriction digestion analysis. The verified clones were designated as pMH5002 (á-RNAP + NifX), pBG1786 (á-RNAP +N-terminal (N1) NifX), pBG1787 (á-RNAP +C-terminal (C2) NifX), pMH6002 (ëCI + NifB), pBG1778 (ëCI + NifN) and pBG1790 (ëCI + NafY). The construction of the plasmids pBG1716 (á-RNAP + NifK), pBG1718 (ëCI + NifK) and pBG1866 (ëCI + NifH) has been described elsewhere (Verma, 2002; Lahiri et al., 2005). Subsequently, desired combinations of plasmids (eg. pTRG-NifX + pBT-NafY) were used for cotransformation

of *E. coli* XL1-Blue cells containing the reporter cassette. The resultant cotransformants were used for performing □-galactosidase activity assays.

â-*Galactosidase assay.* The â-galactosidase activity assay was performed as described under 'Molecular cloning' (Sambrook et al., 1992). Briefly, single *E. coli* transformants were inoculated into 5 ml of 2YT media supplemented with 34 µg/mlof chloramphenicol and 5 µg/mlof tetracycline. The cells were then incubated overnight at 37 °C with shaking at 250 rpm. Then 200 µlof the overnight culture was diluted into 5 ml of the same media and incubated at 37 °C with shaking, until OD_{600} was between 0.3 and 0.6 (initial OD_{600} was 0.05). Cells from 1.5 ml of the culture were collected by centrifugation and resuspended in 500 µlZ-buffer (Sambrook et al., 1992). A 100 µlaliquot of the resuspended cells was lysed by adding 50 µlof chloroform and 25 µlof 0.1% SDS and kept at 28 °C with shaking for 5–10 min. To measure the â-galactosidase activity from the cell lysate, 900 µlof Z-buffer-â-mercaptoethanol solution (0.27 ml of â-mercaptoethanol per 100 ml of Z-buffer) was added to 100 µlculture followed by 200 µl Z-buffer/ONPG (*o*-nitrophenyl-â-D-galactopyranoside—4 mg ONPG per ml of Z-buffer). The time of ONPG addition was recorded, and the tubes were incubated at 37 °C. When yellow color was visible, 500 µlof 1 M Na_2CO_3 was added to each tube to terminate the reaction, and the time was recorded. The optical densities at 420 nm as well as 550 nm were recorded. The â-galactosidase units were defined as the amount of enzyme which hydrolyzes 1 µmolof ONPG to o-nitrophenol and D-galactose per minute. â-Galactosidase activity in Miller Units was calculated as

$1000 \times [OD_{420} - (1.75 \times OD_{550})/(t \times V \times OD_{600})]$, where t is the elapsed time (in min) of

incubation, *V* is the 0.1 ml × concentration factor, and OD_{600} is the absorbance of 1 ml of culture at 600 nm. The results of the □-galactosidase assay were verified for significance by using the student's t-test (GraphPad Software: http://www.graphpad.com/quickcalcs/index.cfm)

Structural Homology Modeling. The Swiss PDB software (freely available online) (Schwede et al., 2003) was used for homology modeling and 3D representation of the derived NifX protein structure. The NafY structure (PDB ID: 1P90 (Dyer et al., 2003)) was used as the template for superimposing the raw NifX amino acid sequence and deriving its 3D structure.

Results And Discussion

Interaction of NifB with NifX. The BacterioMatch™ Two-Hybrid System was used to determine which proteins in the FeMoco maturation pathway directly interacted with NifX. The NifB protein provides an Fe-S precursor, termed NifB-co, for FeMoco assembly (Shah et al., 1994; Allen et al., 1995). Since the *nifB* gene product was found to be essential for the functioning of alternative nitrogenases such as vanadium nitrogenases and iron nitrogenases (Mo-independent) also, therefore it is thought that the NifB-co represents the Fe-S primary cluster for both, the synthesis of FeMoco and the cofactors of alternative nitrogenases (Shah et al., 1994; Joerger and Bishop, 1988). It is believed that the NifB-co is transferred from NifB to NifEN, which acts as a site for rearrangement of the primary [Fe-S] to form the [Fe-S] core of the FeMoco, probably after a structural contact between NifEN and NifH (Allen et al., 1995). The NifB protein sequence comprises of two different domains: an N-terminal SAM radical domain and a C-terminal

NifX-like domain, that is a domain present in certain FeMoco binding proteins (Marchler-Bauer and Bryant, 2004). The presence of the C-terminal NifX-like domain could indicate that this region in NifB is either involved in binding NifB-co precursors and forming the NifB-co or binding a preformed NifB-co for transfer to another protein in the FeMoco biosynthetic pathway (Curatti et al., 2006).

Although it was evident from previous reports that NifX is capable of binding to NifB-co in one of the stages during FeMoco biosynthesis (Rangaraj et al., 2001), a direct protein-protein interaction between the NifB and NifX proteins had not been demonstrated before. Our results support the idea that there is an interaction between NifB and NifX as found by measuring the □-galactosidase activity of *E. coli* cotransformants harboring the plasmids pMH6002 (□CI:NifB) and pMH5002 (RNAP:NifX). As compared to the □-galactosidase activity obtained from *E. coli* cotransformants harboring only the pBT and pTRG plasmids (negative control--23 ± 6.12 Miller Units), the activity obtained from cells harboring pMH6002 and pMH5002 was much higher (50 ± 9.11 Miller Units), indicating a protein-protein interaction between NifB and NifX (Table 4.2). Since the NifB and NafY proteins share many similarities between their amino acid sequences and since NafY was shown to bind to the apodinitrogenase (Dyer et al., 2003), we also examined the ability of NifB and NifK to interact with each other and found that the □-galactosidase activity of *E. coli* cotransformants harboring the plasmids pMH6002 (□CI:NifB) and pBG1716 (RNAP:NifK) was 20 ± 4.78 Miller Units, suggesting lack of interaction between NifB and NifK.

Interaction of NafY with NifX. The *A. vinelandii* NafY protein (nitrogenase accessory factor Y, also known as □) is a molecular chaperone that assists in maintaining the apodinitrogenase in a conformation that facilitates the insertion of the FeMoco by acting as a metallo-insertase (Rubio et al., 2002). In a study that involved utilization of ^{99}Mo and a purified *in vitro* FeMo-co biosynthesis system for investigating the incorporation of the ^{99}Mo radiolabel into proteins involved in the biosynthesis of FeMo-co, it was found that there was transfer of the ^{99}Mo label from NifH and NifX to NafY or apodinitrogenase (Rangaraj and Ludden, 2002).

These observations in combination with our data strengthen the role of NifX as a transporter of a fully formed FeMoco cluster to NafY. We have detected a direct protein-protein interaction between NifX and NafY, as evident from the □-galactosidase activity (42 ± 8.62 Miller Units) of *E. coli* cells harboring the plasmids pMH5002 (RNAP:NifX) and pBG1790 (□CI:NafY) (Table 4.2). Through earlier studies, it was predicted that the NifX was involved in the latter part of the biosynthetic pathway of FeMoco formation (Rangaraj and Ludden, 2002). Our results confirmed the above prediction, because we were able to identify an interaction between NifX with NafY but neither between NifX and NifN or NifX and NifH, both of which are involved in the initial steps of the pathway. The □-galactosidase activity corresponding to the interaction between NifH and NifX was 33 ± 6.67 Miller Units and between NifN and NifX was 31 ± 6.05 Miller Units, indicating absence of any interaction between these pairs of proteins, when compared to the negative and positive controls (23 ± 6.12 and 116 ± 25.01 Miller Units respectively) (Table 4.2).

Table 4.2. Results of the liquid â-galactosidase assay with ONPG as substrate to demonstrate protein-protein interaction

Plasmid to which □CI was translationally fused	Plasmid to which □-RNAP was translationally fused	□-Galactosidase activity (Miller units)[a]	Interacting Peptides
pBG1713 (□CI: NifD-K fusion)	pMH5002 (RNAP: NifX)	45 ± 8.72	NifD-K and Nif X
pMH5007 (□CI: NifD)	pMH5002 (RNAP: NifX)	37 ± 3.74	None
pBG1718 (□CI: NifK)	pMH5002 (RNAP: NifX)	46 ± 8.81	NifK and NifX
pBG1718 (□CI: NifK)	pBG1786 (RNAP:N1NifX)	34 ± 8.74	None
pBG1718 (□CI: NifK)	pBG1787 (RNAP:C2NifX)	35 ± 7.81	None
pBG1778 (□CI: NifN)	pMH5002 (RNAP: NifX)	31 ± 6.05	None
PBG1866 (□CI: NifH)	pMH5002 (RNAP: NifX)	33 ± 6.67	None
pMH6002 (□CI: NifB)	pMH5002 (RNAP: NifX)	50 ± 9.11	NifB and NifX
pMH6002 (□CI: NifB)	pBG1716 (RNAP: NifK)	20 ± 4.78	None
pBG1790 (□CI: NafY)	pMH5002 (RNAP: NifX)	42 ± 8.62	NafY and NifX
pBG1790 (□CI: NafY)	pBG1716 (RNAP: NifK)	56 ± 9.30	NafY and NifK
pBT	pMH5002 (RNAP: NifX)	30 ± 1.63	None
pBT	PTRG	23 ± 6.12	None
pBT-LGF2	pTRG-Gal11^{P}	116 ± 25.01	Mutant form of Gal1 protein with Gal11^{P}

[a]Each assay was performed a minimum of four times for accuracy and the â-galactosidase activity units shown are an average of three independent observations.

We also tested the validity of the results obtained through the BacterioMatch^{TM} Two Hybrid system by determining whether the NafY and NifK proteins, that have been reported to associate with each other based on immunoblot analyses in earlier studies (Dyer et al., 2002), showed evidence of interaction with each other by this method. Our results established that a strong interaction exists between the NafY and NifK proteins (56 ± 9.30 Miller Units) (Table 4.2).

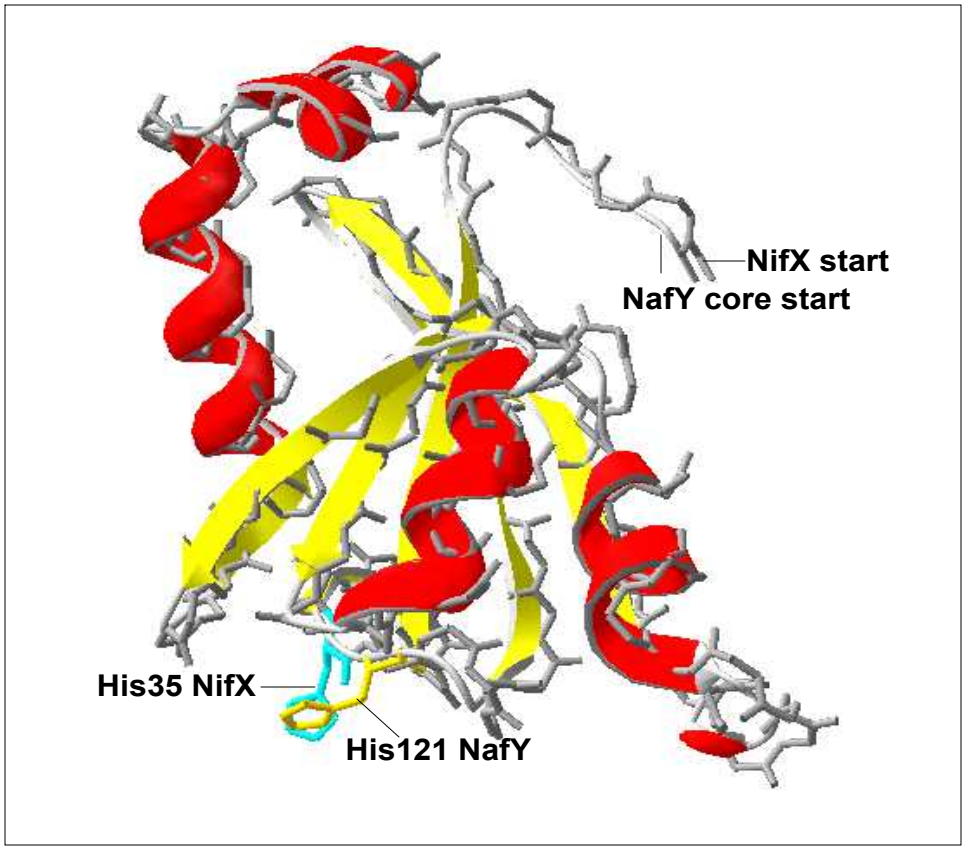
Interaction of NifDK (fusion) and NifK with NifX. A role for NifX in the insertion process of the FeMoco into apodinitrogenase has been suggested earlier (Rangaraj and Ludden, 2002). In that study, the *in vitro* FeMoco biosynthesis assay was performed and the incorporation of ^{99}Mo radiolabel into apodinitrogenase was examined in reactions including and excluding NifX. It was found that the minus NifX reaction showed lower levels of ^{99}Mo label incorporated into dinitrogenase in comparison to the NifX containing reaction (Rangaraj and Ludden, 2002). Evidence obtained through our two-hybrid data clearly indicates the presence of protein-protein interaction between the NifDK protein encoding a single fused unit of dinitrogenase (Suh et al., 2003) and NifX and between the NifK component of the dinitrogenase and NifX. The □-galactosidase activity of *E. coli* cotransformants harboring the plasmids pBG1713 (□CI: NifD-K fusion) and pMH5002 (□-RNAP:NifX) was 45 ± 8.72 Miller Units and that of pBG1718 (□CI: NifK) and pMH5002 (□-RNAP:NifX) was 46 ± 8.81. The □-galactosidase activity of the positive control that included cotransformed plasmids encoding two known interacting proteins – a mutant form of Gal1 protein and the Gal11^{P}, fused to □CI of pBT and □-RNAP of pTRG respectively, was 116 ± 25.01 Miller Units and value for the negative control that included cotransformed plasmids pBT and pTRG only was 23 ± 6.12 Miller Units. The □-galactosidase activity corresponding to interaction between NifD and NifX was 37 ± 3.74 Miller Units, suggesting that there was very weak or no interaction between the two proteins (Table 4.2). To further analyze if the NifX protein possessed specific or separate domains that helped in its interaction with the NifK protein, we cloned the region spanning 1-300 bp of *nifX*, encoding the N-terminal half of the NifX protein (N1X) and

the region spanning 216-477 bp of *nifX*, encoding the C-terminal half of NifX (C2X), into the pTRG plasmid (pBG1786 and pBG1787 respectively). The □-galactosidase activity indicating the extent of interaction between NifK and NifN1X was 34 ± 8.74 Miller Units and between NifK and NifC2X was 35 ± 7.81 Miller Units (Table 4.2). Thus, the N-terminal and C-terminal NifX domains were unable to retain the ability to interact with the NifK protein individually, indicating that either the entire NifX protein is essential for interaction with NifK or a domain comprising an overlapping region from the N-terminal and C-terminal domains contains the necessary stretch of residues for interacting with NifK.

The NifX family of proteins comprises of proteins like NifB, NifY, NafY, VnfX and NifX and is based on the presence of a FeMoco-cluster binding domain. As seen in Fig. 4.1., the NifX protein is largely similar to the C-terminal domain of the NafY protein. The C-terminal NafY protein known as the 'core' domain, was shown to be capable of binding to the FeMoco but unable to bind to the apodinitrogenase in the absence of the first domain, indicating that the N-terminal region of NafY was important in interacting with dinitrogenase (Dyer et al., 2003). Our studies have demonstrated that the NifX is capable of interacting with the NifK protein. The fact that the NifX does not have a domain corresponding to the N-terminal region of NafY and is still able to interact with NifK could be as a result of the conservation of a few residues at the beginning of the NifX protein that are important in maintaining an interaction with NifK (Fig. 4.2). It could also be that whereas NafY interacts with a different region of the dinitrogenase protein, the NifX acts upon a separate region of the same protein and therefore the NifX

protein which could actually be considered as only the C-terminal half of NafY, may be involved in the dual role of FeMoco binding and dinitrogenase binding. Utilizing a simple two-hybrid approach, we have been able to reveal the following important interactions between the NifX protein and some other proteins related to the FeMoco biosynthetic pathway of nitrogenase. Mainly, the following interactions were found: (i) NifB and NifX (ii) NafY and NifX (iii) NifDK (fusion) and NifX (iv) NifK and NifX. Further protein-protein interaction based studies involving NifU, NifS, NifQ and NifV with NifX would certainly prove useful in better deciphering the role of NifX and the events of the FeMoco biosynthetic pathway as a whole.

However, based on our current findings and several previous studies on the FeMoco biosynthetic pathway, we propose a model that best explains the order of events that occurs during the FeMoco biosynthesis, assembly and insertion. It should be noted that due to the complexity of the FeMoco biosynthetic pathway, several different views of the exact sequence of events have been proposed. Therefore, assigning confirmed roles to its participant proteins has been difficult. As seen in our suggested model in Fig. 4.3, the FeMoco biosynthetic pathway starts with the initial mobilization and assembly of the Fe-S fragments and is carried out by the NifUS complex (Yuvaniyama et al., 2000); This Fe-S fragment is then delivered to NifB and a Fe-S core is formed on it (Shah et al., 1994). Following this, the core is transferred from NifB to the NifEN complex.



The NafY core domain structure (PDB ID: 1P90) was used as a template for homology modeling of NifX. The NafY structure is represented in ribbons and the NifX structure in backbone. The NafY core domain start (Glu99) and NifX start (Met 1) are indicated by arrows. The His121 ligand of NafY that is known to be important for FeMoco binding (11) is shown to coincide with the His35 residue of NifX.

Fig. 4.1. Homology modeling of NifX using NafY core structure

Based on previous studies and the detection of an interaction between NifB and NifX in our study, it could be possible that NifX facilitates the transfer of the core from NifB to NifEN (Rangaraj et al., 2001). It is possible that the addition of homocitrate takes place at this step, since one previous study has shown that the NifX may be involved in specifying the organic acid moiety of the FeMoco precursor (Rangaraj and

Ludden, 2002). Following this, multiple FeMoco processing steps take place on the NifEN complex, which acts as a scaffold for this purpose (Dos Santos et al., 2004; Rubio et al., 2005).

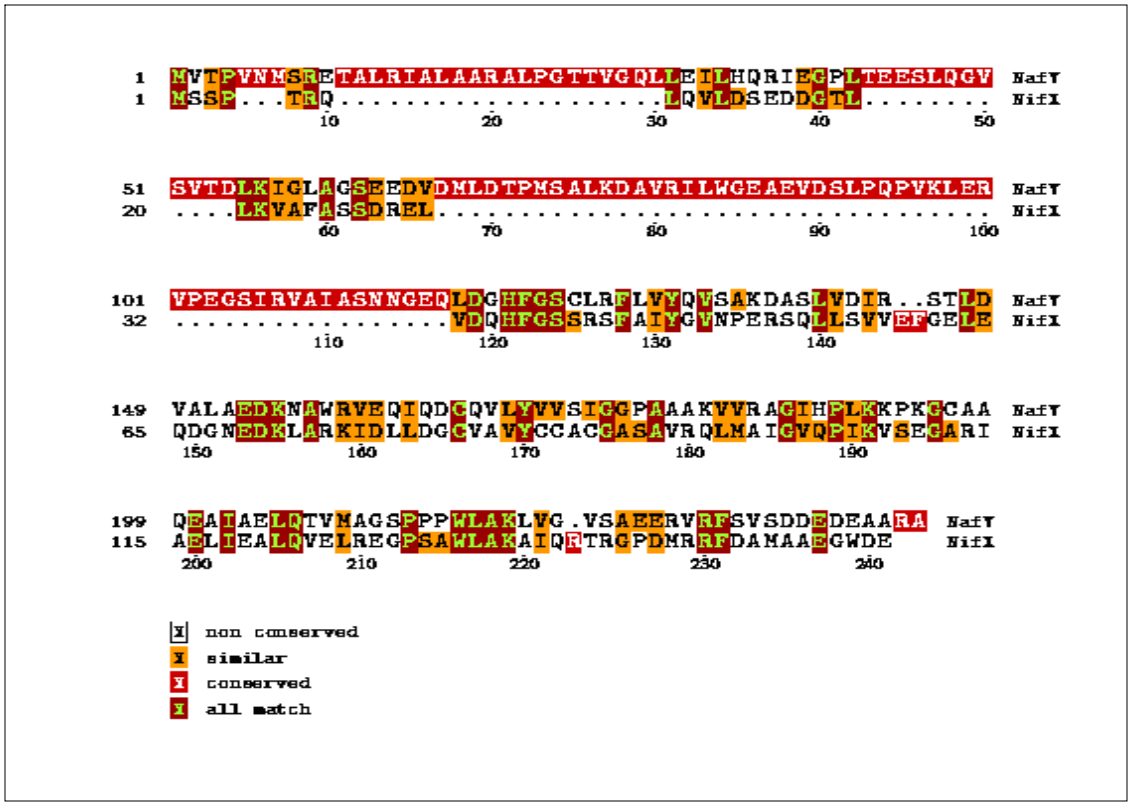
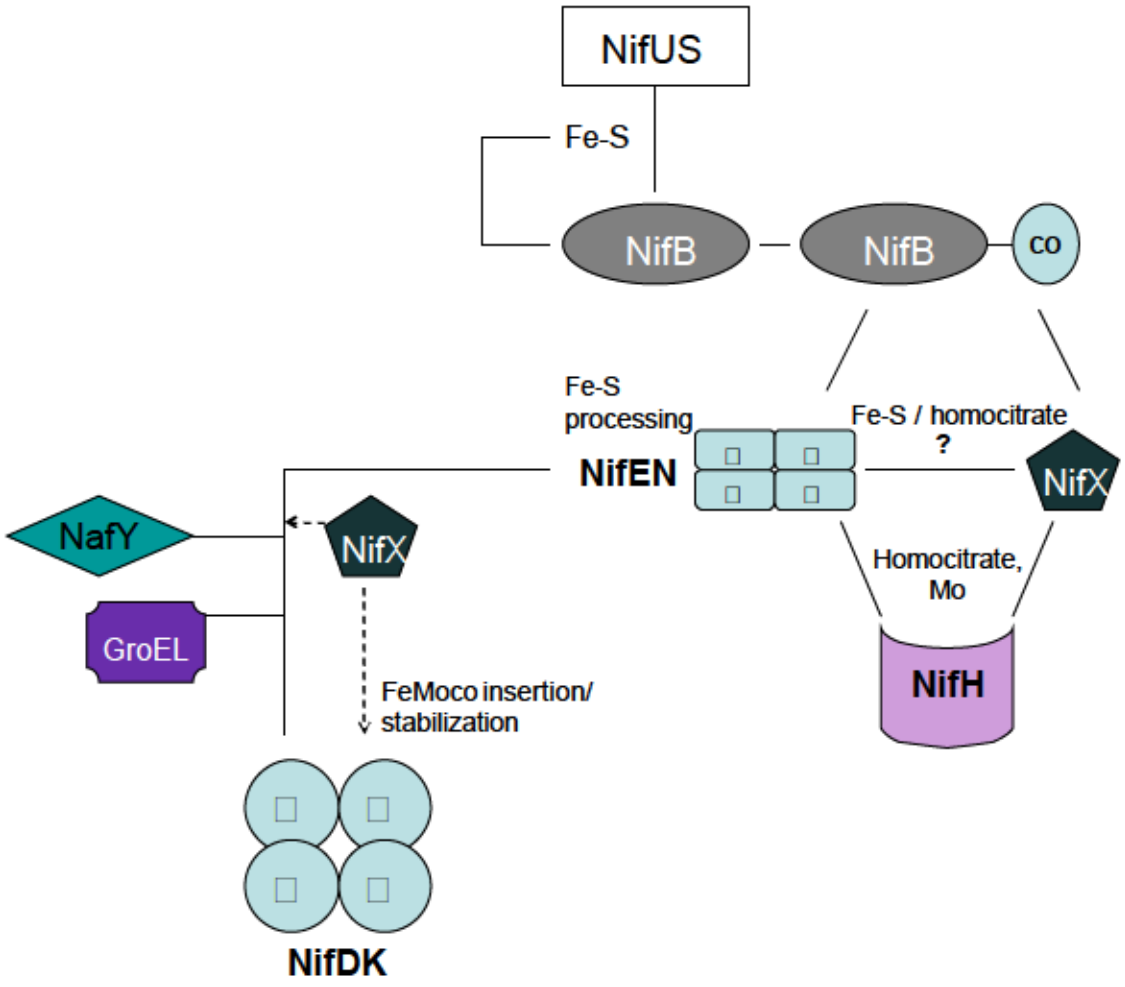


Fig. 4.2. ClustalW alignment of NifX and NafY protein sequences from *A. vinelandii*.

Very recently, Corbett *et. al* showed through x-ray absorption spectroscopy studies that the first isolatable FeMoco precursor on NifEN was a molybdenum-free analog of FeMoco and that the molybdenum was probably incorporated afterwards at the position previously occupied by the non-cysteine ligated terminal iron atom (Corbett et al., 2006). As found in other recent studies, the NifH may perform the role of facilitating

the insertion of Mo and homocitrate into the FeMoco, probably after MgATP hydrolysis (Hu et al., 2005). The GroEL protein is also required for the final assembly of the MoFe protein; It functions as a molecular chaperone for insertion of FeMoco into the MoFe protein (Ribbe and Burgess, 2001). As found in our studies, the interaction of NifX with NafY indicates that the NifX may again function as an escort protein at this step and participate in the transfer of the intermediate or completed FeMoco cluster to the NafY protein. In the following steps, the assembled FeMoco that has been delivered to the NafY protein is carried to the apo-dinitrogenase. It has been suggested that the NafY protein dissociates from the MoFe protein after the final insertion process in completed. Since we have detected an interaction between the NifX and NifK proteins in this work, we propose that the NifX may perform a similar role to that of NafY and the final insertion of the FeMoco into the apo-dinitrogenase may involve the participation of NifX. It may also be possible that the NifX stabilizes the apo-dinitrogenase in a conformation suitable for FeMoco insertion through its interaction with NifK. There have been earlier speculations concerning the overlapping or redundant roles of the proteins involved in the FeMoco biosynthetic pathway, for example, the role of NifY was found to be similar to that of NafY in studies involving the FeMoco biosynthetic pathway in *K. pneumoniae* and *A. vinelandii* (Homer et al., 1993). However, this occurrence has not been yet explainable and future studies that investigate the FeMoco biosynthetic pathway in detail may lead to meaningful answers.



The NifU and NifS proteins participate in the initial Fe/S mobilization; this precursor Fe-S moiety is then transferred to the NifB protein, which is then termed as NifB-co. The NifX protein may play a role as a molybdenum (Mo) carrier from the NifB to the NifEN complex or it may provide the homocitrate moiety to NifEN. This processed FeMoco is inserted and stabilized in the NifDK (MoFe) protein and this activity is assisted by NifX, NafY and GroEL.

Fig. 4.3. Proposed model for the role of NifX in FeMoco biosynthesis/insertion pathway.

References

Allen, R. M., Chatterjee, R. Ludden, P. W. and Shah, V. K. 1996. The requirement of reductant for *in vitro* biosynthesis of the iron-molybdenum cofactor of nitrogenase. J. Biol. Chem. 271: 4256-4260.

Allen, R. M., Chatterjee, R., Ludden, P. W. & Shah, V. K. 1995. Incorporation of iron and sulfur from NifB cofactor into the iron-molybdenum cofactor of dinitrogenase. J. Biol. Chem. 270: 26890-26896.

Anantharaman, V., Koonin, E. V. and Aravind, L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J. Mol. Biol. 307(5): 1271-1292.

BacterioMatch Two-Hybrid System Instruction Manual, 2001, Stratagene, Inc., La Jolla, CA.

Barney, B. M., Lee, H. I., Dos Santos, P. C., Hoffman, B. M., Dean, D. R. and Seefeldt, S. C. 2006. Breaking the N_2 triple bond: insights into the nitrogenase mechanism. Dalton Trans. (19): 2277-84. Epub 2006 Apr 11.

Chan, M. K., Kim, J. and Rees, D. C. 1993. The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 A resolution structures. Science 260: 792–794.

Chatterjee, R., Allen, R. M., Shah, V. K. and Ludden, P. W. 1994. Nucleotide and divalent cation specificity of *in vitro* iron-molybdenum cofactor synthesis. J. Bacteriol. 176: 2747-2750.

Corbett, M. C., Hu, Y., Fay, A. W., Ribbe, M. W., Hedman, B. and Hodgson, K. O. 2006. Structural insights into a protein-bound iron-molybdenum cofactor precursor. Proc. Natl. Acad. Sci. U S A 103(5): 1238–1243.

Curatti, L., Ludden, P. W. and Rubio, L. M. 2006. NifB-dependent *in vitro* synthesis of the iron–molybdenum cofactor of nitrogenase. Proc. Natl. Acad. Sci. U S A 103(14): 5297–5301.

Dos Santos, P. C., Dean, D. R., Hu, Y. and Ribbe, M. W. 2004 Formation and insertion of the nitrogenase iron-molybdenum cofactor. Chem. Rev. 104: 1159–1174

Dove, S. L., Joung, J. K. and Hochschild, J. K. A. 1997. Activation of prokaryotic transcription through arbitrary protein–protein contacts. Nature 386: 627–630

Dyer, D. H., Rubio, L. M., Thoden, J. B., Holden, H. M., Ludden, P. W. and Rayment, I. 2003. The three-dimensional structure of the core domain of NafY from *Azotobacter vinelandii* determined at 1.8-A resolution. 278(34): 32150-6.

Einsle, O., Tezcan, F. A., Andrade, S. L., Schmid, B., Yoshida, M., Howard, J. B. and Rees, D. C. 2002. Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor. Science 297: 1696–1700.

Goodwin, P. J., Agar, J. N., Roll, J. T., Roberts, G. P., Johnson, M. K. and Dean, D. R. 1998. The *Azotobacter vinelandii* NifEN complex contains two identical [4Fe-4S] clusters. Biochemistry 37: 10420-10428.

Homer, M. J., Dean, D. R. and Roberts, G. P. 1995. Characterization of the ã protein and its involvement in the metallocluster assembly and maturation of dinitrogenase from *Azotobacter vinelandii*. J. Biol. Chem. 270: 24745-24752.

Homer, M. J., Paustian, T. D., Shah, V. K. and Roberts, G. P. 1993. The *nifY* product of *Klebsiella pneumoniae* is associated with apodinitrogenase and dissociates upon activation with the iron-molybdenum cofactor. J. Bacteriol. 175(15): 4907–4910.

Howard, J. B. and Rees, D. C. 1996. Structural basis of biological nitrogen fixation, Chem. Rev. 96: 2965–2982.

Hu, Y., Fay, A. W. and Ribbe, M. W. 2005. Identification of a nitrogenase FeMo cofactor precursor on NifEN complex. Proc. Natl. Acad. Sci. U S A 102(9): 3236–3241.

Igarashi, R. Y. and Seefeldt, L. C. 2003. Nitrogen fixation: the mechanism of the Mo-dependent nitrogenase. Crit. Rev. Biochem. Mol. Biol. 38: 351–384.

Imperial, J., Shah, V. K., Ugalde, R. A., Ludden, P. W., Brill, W. J. 1987. Iron-molybdenum cofactor synthesis in *Azotobacter vinelandii nif*⁻ mutants. J. Bacteriol. 169: 1784–1786.

Imperial, J., Ugalde, R. A., Shah, V. K., and Brill, W. J. 1984. Role of the *nifQ* gene product in the incorporation of molybdenum into nitrogenase in *Klebsiella pneumoniae*. J. Bacteriol. 158: 187-94

Jacobson, M.R., Brigle, K. E., Bennett, L. T., Setterquist, R. A.,Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E. and Dean, D. R. 1989. Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. J. Bacteriol. 171(2): 1017-27.

Joerger, R. D. and Bishop, P. E. 1988. Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. J. Bacteriol. 170: 1475-1487.

Kennedy, C. and Dean, D. R. 1992. The *nifU*, *nifS* and *nifV* gene products are required for activity of all three nitrogenases of *Azotobacter vinelandii*. Mol. Gen. Genet. 231: 494-8.

Lahiri. S., Pulakat, L. and Gavini, N. 2005. Functional NifD-K fusion protein in *Azotobacter vinelandii* is a homodimeric complex equivalent to the native heterotetrameric MoFe protein. Biochem. Biophys. Res. Commun. 337(2): 677-684

Marchler-Bauer, A., Bryant, S. H. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32(W): 327-331

Meijer, W. G. and Tabita, F. R. 1992. Isolation and characterization of the *nifUSVW-rpoN* gene cluster from *Rhodobacter sphaeroides*. J. Bacteriol. 174: 3855-3866.

Paustian, T. D., Shah, V. K. and Roberts, G. P. 1989. Purification and characterization of the *nifN* and *nifE* gene products from *Azotobacter vinelandii* mutant UW45. Proc. Natl. Acad. Sci. USA 86: 6082-6086.

Peters, J. W., Fisher, K. and Dean, D. R. 1995. Nitrogenase structure and function: a biochemical-genetic perspective. Annu. Rev. Microbiol. 49: 335–366.

Rangaraj, P. and Ludden, P. W. 2002. Accumulation of ^{99}Mo-containing Iron-Molybdenum Cofactor Precursors of Nitrogenase on NifNE, NifH, and NifX of *Azotobacter vinelandii*. J. Biol. Chem. 277(42): 40106-40111

Rangaraj, P., Ruttimann-Johnson, C., Shah, V. K. and Ludden, P. W. 2001. Accumulation of ^{55}Fe-labeled precursors of the iron-molybdenum cofactor of nitrogenase on NifH and NifX of *Azotobacter vinelandii*. J. Biol. Chem. 276(19): 15968-15974

Rees, D. C., Tezcan, F. A., Haynes, C. A., Walton, M. Y., Andrade, S., Einsle, O. and Howard, J. B. 2005. Structural basis of biological nitrogen fixation. Philos. Trans. R. Soc. London A 363: 971–984.

Ribbe, M. W. and Burgess, B. K. 2001. The chaperone GroEL is required for the final assembly of the molybdenum-iron protein of nitrogenase. Proc. Natl. Acad. Sci. USA 98(10): 5521-5525

Robinson, A. C., Dean, D. R. and Burgess, B. K. 1987. Ironmolybdenum cofactor biosynthesis in *Azotobacter vinelandii* requires the iron protein of nitrogenase. J. Biol. Chem. 262:14327-14332.

Roll, J. T., Shah, V. K., Dean, D. R. and Roberts, G. P. 1995. Characteristics of NIFNE in *Azotobacter vinelandii* strains. Implications for the synthesis of the iron-molybdenum cofactor of dinitrogenase. J. Biol. Chem. 270: 4432-4437.

Rubio, L. M. and Ludden, P. W. 2005. Maturation of Nitrogenase: a Biochemical Puzzle. J. Bacteriol. 187: 405–414.

Rubio, L. M., Rangaraj, P., Homer, M. J., Roberts, G. P. & Ludden, P. W. 2002. Cloning and mutational analysis of the ã gene from *Azotobacter vinelandii* defines a new family of proteins capable of metallocluster binding and protein stabilization. J. Biol. Chem. 277: 14299-14305.

Rubio, L. M., Singer, S. W. and Ludden, P. W. 2004. Purification and characterization of NafY (apodinitrogenase {gamma} subunit) from *Azotobacter vinelandii*. J. Biol. Chem. 279(19): 19739-19746.

Sambrook, J. F. Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1992

Schwede, T., Kopp, J. Guex, N. and Peitsch, M. C. 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucl. Acids Res. 31: 3381-3385

Shah, V. K., Allen, J. R., Spangler, N. J. and Ludden, P. W. 1994. *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. Purification and characterization of NifB cofactor, the product of NIFB protein. J. Biol. Chem. 269: 1154-1158.

Shah, V. K., Imperial, J., Ugalde, R. A., Ludden, P. W. and Brill, W. J. 1986. *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. Proc. Natl. Acad. Sci. USA 83: 1636-1640.

Shah, V. K., Rangaraj, P., Chatterjee, R., Allen, R. M., Roll, J. T., Roberts, G. P. and Ludden, P. W. 1999. Requirement of NifX and other Nif proteins for *in vitro* biosynthesis of the iron-molybdenum cofactor of nitrogenase. J. Bacteriol. 181(9): 2797–2801.

Suh, M-H., Pulakat, L. and Gavini, N. 2003. Functional expression of a fusion-dimeric MoFe protein of nitrogenase in *Azotobacter vinelandii*. 278(7): 5353-60.

Ugalde, R. A., Imperial, J., Shah, V. K., Brill, W. J. 1984. Biosynthesis of iron-molybdenum cofactor in the absence of nitrogenase. J. Bacteriol. 159: 888–893

Verma, S. 2002. Investigation of the NifM-NifH interaction using BacterioMatch Two Hybrid System. Masters thesis. Biological Sciences Department, Bowling Green State University, Ohio, USA.

Yuvaniyama, P., Agar, J. N., Cash, V. L., Johnson, M. K. and Dean, D. R. 2000. NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein. Proc. Natl. Acad. Sci. U.S.A. 97: 599-604.

Zheng, L. M., White, R. H. and Dean, D. R. 1997. Purification of the *Azotobacter vinelandii nifV*-encoded homocitrate synthase. J. Bacteriol. 179: 5963-5966.

CHAPTER V

NIFH: STRUCTURAL AND MECHANISTIC SIMILARITIES WITH PROTEINS INVOLVED IN DIVERSE BIOLOGICAL PROCESSES

Introduction

The nitrogenase enzyme is a two-component system that consists of the iron protein (Fe-protein or NifH) and molybdenum-iron protein (MoFe-protein) working in concert to effect nitrogen reduction (*for reviews, see:* Rees et al., 2005; Rees, 2002; Seefeldt and Dean, 1997; Burgess and Lowe, 1996). The 64 kDa homodimeric NifH protein has two ATP binding domains and one [4Fe-4S] cluster per homodimer. It supplies energy by ATP hydrolysis and transfers electrons from reduced ferredoxin or flavodoxin to the MoFe-protein (Seefeldt and Dean, 1997). It can bind up to two MgADP and/or MgATP molecules per dimer and is essential for coupling nucleotide hydrolysis to electron transfer for eventual nitrogen reduction (Georgiadis et al., 1992; Chen et al., 1994). It also participates in the biosynthesis and insertion of the FeMo-cofactor into the MoFe protein (Tal et al., 1991; Gavini and Burgess, 1992, Ma et al., 1994; Gavini et al., 1994). In the past, various proteins have been revealed to have structural and mechanistic similarities as well as evolutionary relationships with the NifH protein, notable among them being: light independent protochlorophyllide (Pchlide) reductase (ChlL/FrxC or bChL), arsenite pump ATPase (ArsA), 2-hydroxyglutaryl

dehydratase Component A (CompA) involved in glutamate degradation and MinD that functions in spatial regulation of cell division (Lutkenhaus and Sundaramoorthy, 2003; Hu et al., 2003; Gatti et al., 2000; Fujita and Bauer, 2000; de Boer et al., 1992; Fujita et al., 1989). Although involved in diverse biological processes, these proteins have been found to bear considerable structural resemblance to the NifH protein. Mechanistically, they do play a role similar to NifH in their respective complexes, based upon the presence of ATP binding motifs in each of them (Schmid et al., 2002; Hayashi et al., 200; Locher et al., 2001; Zhou et al., 2000; Koonin, 1993). Whereas ChlL, ArsA and MinD belong to the same superfamily of proteins as NifH, known as the 'P-loop containing nucleoside triphosphate hydrolases', the Component A of 2-hydroxyglutaryl dehydratase (CompA) is a member of the 'Actin-like ATPase domain' superfamily, as designated by the SCOP database (Murzin et al., 1995). An analogy between NifH and ChlL, ArsA, CompA and MinD lies in the organization of these proteins as members of their respective two component systems. In the nitrogenase complex, NifH functions as the obligate electron donor to its specific partner, the MoFe protein. This intermolecular electron transfer process requires ATP hydrolysis. In a similar manner, for the dehydratase system, CompA is the site of ATP hydrolysis and is the electron donor to the second component, CompD (Locher et al., 2001); ArsA exists in a complex with the ArsB protein forming an arsenite-antimonite [As(III)/Sb(III)]- translocating ATPase. As a result of the ATP hydrolysis in ArsA, a more compact conformation of the enzyme is generated, allowing the vectorial movement of the arsenite or antimonite ions into ArsB (Gatti et al., 2000); It has been speculated that the MinD-MinE membrane complex is analogous to the NifH-

MoFe complex and results in ATP hydrolysis, as a consequence of which, MinD and MinE are released from the membrane (Lutkenhaus and Sundaramoorthy, 2003); Comparison of the molecular architecture between nitrogenase and the light-independent Pchlide reductase has shown that the NifH counterpart in the light independent Pchlide reductase is ChlL (in *Chlamydomonas reinhardtii*). ChlL is suggested to be involved in ATP-dependent transfer of electrons from a reductant, such as ferredoxin, to the ChlB-ChlN complex via the Fe:S center (Fujita and Bauer, 2000). Thus, in a manner similar to NifH, the proteins ArsA, ChlL, CompA and MinD also couple hydrolysis of nucleoside triphosphates to redox reactions from a metallic cluster.

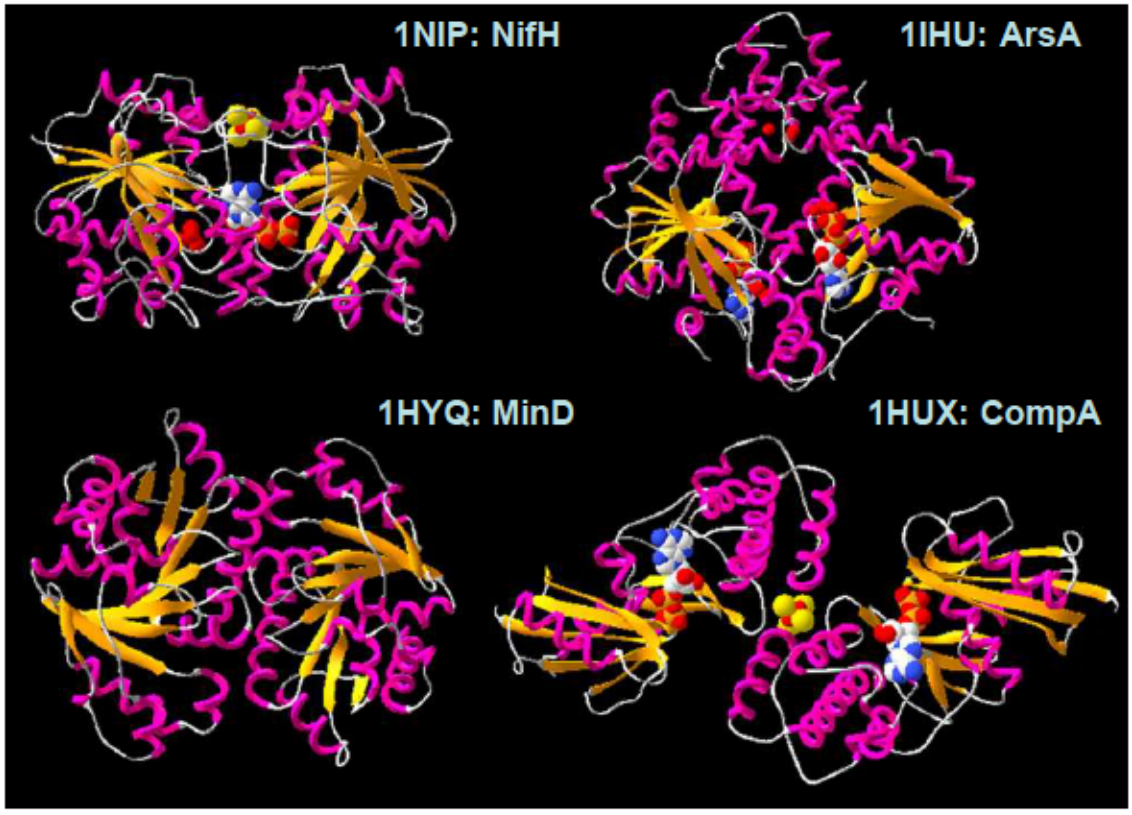
The protein sequence comparison of NifH, ArsA, ChlL, CompA and MinD gave a broad view of their similar regions (Fig. 5.1). However, it is difficult to detect protein relationships and the presence of different domains by direct comparisons of sequences belonging to different protein families and it is more reasonable to determine the presence or absence of domains and their family relationships by comparison of their three-dimensional structures (Fig. 5.2). How proteins with similar protein folding patterns perform varied functions seems to be a prevalent question and it has been suggested that " 'popular' folding patterns have thermodynamic advantages enabling them to be stabilized by random sequences and these few advantageous folds can probably tolerate various primary structures and therefore perform different functions" (Finkelstein et al., 1993). The structural comparison of each of the four proteins with NifH creates a general model that indicates their common ancestry at some point. The functions of NifH, ArsA and ChlL, which are broadly nitrogen fixation, arsenite resistance and

```
ChlL_ChlRe      ----------------------------------MKLAVYGKGGIGKSTTSCNISIALR  25
NifH_Av         --------------------------------MAMRQCAIYGKGGIGKSTTTQNLVAALA  28
MinD_EColi      --------------------------------MARIIVVTSGKGGVGKTTSSAAIATGLA  28
ArsA1           --------------------------MQFLQNIPPYLFFTGKGGVGKTSISCATAIRLA  33
ArsA2           STQPVASPSSDEYLQQRPDIPSLSALVDDIARNEHGLIMLMGKGGVGKTTMAAAIAVRLA 352
CompA           ----------------------------------MSIYTLG-IDVGSTASKCIILKDGK  24

ChlL_ChlRe      KRGKKVLQIGCDPKHDSTFTLTGFLIPTIIDTLSSKDYHYEDIWPEDVIYGGYGGVDCVE  85
NifH_Av         EMGKKVMIVGCDPKADSTRLILHSKAQNTIMEMAAEAGTVEDLELEDVLKAGYGGVKCVE  88
MinD_EColi      QKGKKTVVIDFDIGLRNLDLIMGCERR-VVYDFVNVIQGDATLNQALIKDKRTENLYILP  87
ArsA1           EQGKRVLLVSTDPASN-VGQVFSQTIGITIQAIASVPGLSALEIDPQAAAQQYRARIVDP  92
ArsA2           DMGFDVHLTTSDPAAHLSMTLN-----------GSLNNLQVSRIDPHEETERYRQHVLET 401
CompA           EIVAKSLVAVGTGTSGPARSIS---------EVLENAHMKKEDMAFTLATGYGRNSLEGI  75

ChlL_ChlRe      AGGPPAGAGCGGYVVGETVKLLKELNAFF-EYDVILFDVLGDVVCGGFAAPLN--YADYC 142
NifH_Av         SGGPEPGVGCAGRGVITAINFLEEEGAYEDDLDFVFYDVLGDVVCGGFAMPIRENKAQEI 148
MinD_EColi      ASQTRDKDALTREGVAKVLDDLKAM-----DFEFIVCDSPAGIETG---ALMALYFADEA 139
ArsA1           IKGVLPDDVVSSINEQLSGACTTEIAAFD-EFTGLLTDASLLTRFDHIIFDTAPTGHTIR 151
ArsA2           KGKELDEAGKRLLEEDLRSPCTEEIAVFQ-AFSRVIREAG----KRFVVMDTAPTGHTLL 456
CompA           ADKQMSELSCHAMGASFIWPNVHTVIDIG-GQDVKVIHVENGTMTN-FQMNDKCAAGTGR 133

ChlL_ChlRe      IIVTDNGFDALFAAN-RIAASVREKARTHPLRLAGLIGN-RTSKRD--LIDKYVEACPMP 198
NifH_Av         YIVCSGEMMAMYAAN-NISKGIVKYANSGSVRLGGLICNSRNTDREDELIIALANKLGTQ 207
MinD_EColi      IITTNPEVSSVRDSD-RILGILASKSRR--------AENGEEPIKEHLLLTRYNPGR-VS 189
ArsA1           LLQLPGAWSSFIDSNPEGASCLGPMAGLEKQREQYAYAVEALSDPKRTRLVLVARLQKST 211
ArsA2           LLDATGAYHREIAKKMGEKGHFTT-------------PMMLLQDPERTKVLLVTLPETTP 503
CompA           FLDVMANILEVKVSDLAELGAKSTKRVAIS------STCTVFAESEVISQLSKGTDKIDI 187

ChlL_ChlRe      VLEVLPLIEEIRISRVKGKTLFEMSNKNNMTSAH--MDGSKGDNSTVGVSETPSEDYICN 256
NifH_Av           MIHFVPRDNVVQRAEIRRMTVIEYDPKAKQADEYRALARKVVDNKLLVIPNPITMDELEE 267
MinD_EColi      RGDMLSMEDVLEILRIKLVGVIPEDQSVLRASNQ-GEPVILDINADAGKAYADTVERLLG 248
ArsA1             LQEVARTHLELAAIGLKNQYLVINGVLPKTEAANDTLAAAIWEREQEALANLPADLAGLP 271
ArsA2           VLEAANLQADLERAGIHPWGWIINNSLSIADTRS----PLLRMRAQQELPQIESVKRQHA 559
CompA           IAGIHRSVASRVIGLANRVGIVKDVVMTGGVAQNYGVRGALEEGLGVEIKTSPLAQYNGA 247

ChlL_ChlRe      FYLNIADQLLTEPEGVIPRELADKELFTLLSDFYLKI 293
NifH_Av         LLMEFG-IMEVEDESIVGKTAEEV------------- 290
MinD_EColi      EERPFR-FIEEEKKGFLKRLFGG-------------- 270
ArsA1           TDTLFLQPVNMVGVSALSRLL---------------- 292
ArsA2           SRVALVPVLASEPTGIDKLKQLAG------------- 583
CompA           LGAALYAYKKAAKSAWSHPQFEK-------------- 270
```

The residues are colored as red, blue, magenta, green and gray based on their hydrophobicity and other properties. Structural regions highlighted are the nucleotide binding domains in gray; the Switch I regions in yellow; the Switch II regions in cyan. Cysteine residues highlighted in red indicate the ligands for the metal centers of their respective proteins. Histidine (148, 453) and serine (420) of the ArsA A1 and ArsA A2 polypeptides are highlighted in red as other ligands for the Sb(III) center of ArsA. It should be noted that ArsA A2 sequence starts after residue 292 of ArsA A1, includes the 'linker' stretch of residues and therefore its numbering starts from residue 293.

Fig. 5.1. Protein sequence alignment of NifH, ChlL, ArsA A1, ArsA A2, MinD and CompA.

photosynthesis respectively, further point out to the very first evolutionarily important processes on earth. It has been speculated that the global functions of the NifH homologs may have varied with either change in nucleotide specificity, the incorporation of new functions or the loss of others (Schlessman et al., 1998).



The structures are colored on the basis of secondary structures. The □-helices are shown in magenta, â-strands in orange and coils/loops are in gray. The yellow spacefilled structure in NifH and CompA represent the [4Fe4S] cluster. The red spacefilled structures in ArsA show the Sb(III)/As(III) metal center. CPK colored spacefilled structures in NifH, ArsA and CompA indicate the bound MgADP. (1HUX: Locher et al., 2001; 1NIP: Georgiadis et al., 1992; 1HYQ: Cordell and Lowe, 2001; 1IHU: Zhou et al., 2001).

Fig. 5.2. 3-D representation of NifH, ArsA, MinD and CompA

A thorough analysis of the comparative structural details of these proteins allows a better understanding of their relationship to each other. Consequently, derived from the comparison of NifH with ChlL, ArsA, CompA and MinD, greater insight into their structure based function can be obtained. For a thorough understanding of the structural and mechanistic features of the NifH protein, its features have been reviewed here.

The 2.9 Å crystal structure of the NifH protein from *Azotobacter vinelandii* was obtained by Georgiadis *et al*. in 1992 (Georgiadis et al., 1992). It revealed that each of the subunits of the NifH dimer consisted of mixed □-helix/□-sheet polypeptide fold, with a consensus topology of an eight-stranded □-sheet flanked by nine □-helices. A □-sheet core in each monomer is formed of one short antiparallel and seven parallel □-strands. The binding sites critical to NifH function were determined by loops at the carboxy-terminal ends of several □-strands. The nucleotide binding sites (NBSs) were situated in the loops following □1 and □2 and the cysteine ligands (C97 and C132) to the [4Fe-4S] cluster were in the loops following □5 and □6 (Georgiadis et al., 1992; Hausinger and Howard, 1983). Both cluster ligands are located near the amino terminal ends of □-helices that are directed toward the cluster. As indicated by spectroscopic studies, a prominent feature of the Fe protein is the exposure of the [4Fe-4S] cluster to the solvent (Howard and Rees, 1996), which is also probably one of the reasons that it is an oxygen sensitive protein. In addition to the [4Fe-4S] cluster, there are numerous Van der Waals and polar interactions in the interface beneath the cluster that help stabilize the dimer structure. NifH displays structural similarity to other nucleotide-binding proteins including signal transduction molecules such as G-proteins and *ras* and energy

transduction systems such as myosin (Seefeldt and Mortenson, 1993). It is known to be similar to the nucleotide binding proteins based on the presence of the following structural features (Milner-White et al., 1991; Schulz, 1992; Sprang, 1997): (i) mainly parallel □-sheets flanked by □-helices (ii) a phosphate-binding loop (P-loop) or Walker A motif (Walker et al., 1982), containing the G-X- X- X- X-G-K-S/T consensus sequence and (iii) two switch regions, Switch I and Switch II that interact with the □-phosphate group of the bound nucleoside triphosphate. Analysis of the position of a bound ADP molecule in the X-ray structure of the Fe protein from *Azotobacter vinelandii* (Georgiadis et al., 1992), and the properties of other site-specifically altered Fe proteins (Gavini and Burgess, 1992; Seefeldt and Mortenson, 1993; Seefeldt et al., 1992; Wolle et al., 1992; Ryle et al., 1995; Lanzilotta et al., 1995), has established the position of the nucleotide binding sites, one on each subunit, some 19 Å away from the [4Fe-4S] cluster. The nucleotide-dependent switch regions are responsible for communication between the sites responsible for nucleotide binding and hydrolysis and the [4Fe-4S] cluster of the Fe protein and the docking interface that interacts with the MoFe protein upon macromolecular complex formation. In *A. vinelandii* NifH, residues 38 to 43 and 125 to 132 form the Switch I and Switch II respectively. The switch regions undergo conformational changes upon hydrolysis of the nucleoside diphosphate with consequent loss of the interaction with the terminal phosphate group. Substitutions in the nucleotide dependent switch regions of the NifH of *Azotobacter vinelandii* showed that the altered NifH proteins formed a trapped complex subsequent to a single electron transfer event (Jang et al., 2004). The studies suggested that whereas in the structure of the native

enzyme the analogous interaction between the side chains of D39 and D125 was precluded due to electrostatic repulsion, the D39N substitution allowed the formation of a hydrogen bond between the Switch I D39 and the Switch II D125; this demonstrated that the electrostatic repulsion between D39 and D125 was important for dissociation of the Fe protein: MoFe protein complex during catalysis (Jang et al., 2004). Thus, the switch regions play a critical role in transmitting information concerning the nucleotide state to other effector molecules that bind to these regions.

Investigations of the dimer interface of the NifH protein have shown that in the NifH dimer, residues with >30Å^2 of buried surface area include K41, E92, P93, V95, A98, D129, V130, V131, C132, M156, Y159, K166 and K170. These residues, along with other residues adjacent in the sequence, mediate the subunit-subunit interactions through a series of primarily polar (hydrogen bond and salt-bridge) interactions (Schlessman et al., 1998). The C-terminus of NifH has been speculated to provide an additional degree of intersubunit interaction in NifH as residues from this region wrap around the body of the opposing subunit and enhance the overall stabilization of the NifH dimer. Also, in the C-terminus region, cross-subunit salt-bridges have been known to form between K224-E277 and K233- E287 (Schlessman et al., 1998).

An integral part of the nitrogenase mechanism is MgATP binding to the Fe protein and hydrolysis by the Fe protein-MoFe protein complex. Through the studies of Jang *et al*., it was confirmed that the protein interactions with the Mg2+ were essential to the transduction of the nucleotide hydrolysis event (Jang et al., 2000). Analysis of the MgADP bound crystallized Fe protein revealed that S16 binds the Mg^{2+} and □, □, and □

phosphates of nucleotides and appears to form a hydrogen bond with D125 of the switch II region (Jang et al., 2000; Seefeldt and Mortenson, 1993). Yet another region found to be critical for the MgATP-induced conformational change was the highly conserved span of the Fe protein around A157 located at the subunit interface. This region is part of the □5 helix extending from residue 151 to 176 at the subunit interface and shows substantial difference when the free Fe protein structure is compared with the structure of the Fe protein in the MgATP bound complex. Moreover, the mutation of A157 to Ser resulted in a protein that could still bind MgATP normally but was unable to undergo the MgATP-induced conformational change (Gavini and Burgess, 1992). Further insight into MgATP binding and hydrolysis was obtained by studies based on residues K15 and D125. It was suggested that the breaking of a salt bridge between these two residues by MgATP binding was responsible for triggering a conformational change (Wolle et al., 1992; Ryle et al., 1995).

The substrate reduction mechanism of the nitrogenase enzyme involves the key step of MoFe protein docking on the Fe protein. Recent studies dealing with the biochemical and structural characterization of the crosslinked complex of nitrogenase have shown the specificity of the residues involved in the transient complex formation between the Fe and MoFe proteins. It was found that only E112 from one of the two NifH subunits and K400 of the NifK subunit were crosslinked, although these residues were surrounded by numerous other charged residues that could potentially participate in this process (Schmid et al., 2002). Most recently, when the Fe and MoFe proteins were cocrystallized under the conditions of (i) no nucleotide (ii) MgADP and (iii)

MgAMPPCP (an MgATP analog), the Fe protein molecules were found to be capable of occupying different interaction sites on the MoFe protein (Tezcan et al., 2005); three separate docking areas were thereby identified on the surface of the MoFe protein (Tezcan et al., 2005). Other studies that substituted Ala in place of F125 of the □ and □ subunits of the MoFe protein, separately and in combination, showed that the doubly substituted MoFe protein was unable to form a tight complex with the MgADP-AlF_4^- treated NifH or when using the altered L127 □NifH, thereby suggesting that the F125 residues were involved in an early event(s) that occurred upon component protein docking and could be involved in eliciting MgATP hydrolysis (Christiansen et al., 2000).

As described before, proteins structurally homologous to NifH include the ChlL, ArsA, CompA and MinD proteins. We investigated the homology modeled structure of ChlL based on the NifH structural template and the structurally aligned CompA, ArsA and MinD structures with respect to the NifH structure in order to detect their detailed structural similarities and dissimilarities. Functionally, all these proteins are involved in diverse roles in nature and therefore the functional complementation of one of these proteins by NifH would indicate the strong conservation of essential domains between the respective proteins. We investigated the genetic complementation of the ArsA protein function by creating a NifH-ArsA2 chimeric protein. The N-terminal (ArsA1) and C-terminal (ArsA2) halves of the ArsA protein are each equivalent to one monomer of the NifH protein. In the NifH-ArsA2 chimeric protein, the N-terminal half of the ArsA was replaced by NifH. We found that the NifH-ArsA2 protein was able to function in arsenite resistance. We further performed phylogenetic analysis using NifH and ArsA protein

sequences from various organisms to investigate their evolutionary relationship and derived important conclusions.

Materials and Methods

Multiple Alignments—Multiple alignments for the amino acid sequences of NifH (*Azotobacter vinelandii*), ArsA1 (plasmid R773 *Escherichia coli*), ArsA2 (plasmid R773 *Escherichia coli*), CompA (*Acidaminococcus fermentas*), MinD (*Escherichia coli*) and ChlL (*Chlamydomonas reinhardtii*) were analyzed by ClustalW (Thompson et al., 1994).

*Structural homology modeling***.** Homology models of ChlL were generated based on the available structural model of the NifH protein (Jang et al., 2000). The Swiss-PDB software (Guex and Peitsch, 1997) was used for performing the homology modeling in this study. We selected a strategy of semi-automatic and manual adjustments of the derived superimposed proteins relative to the available backbone conformation of the NifH protein structure. The structural alignment of the superimposed ChlL protein was done on basis of its CLUSTAL W sequence alignment with NifH. Finally, the images obtained were represented in a graphically refined format using the POV-Ray Imaging software (www.povray.org). Structural superimposition of MinD on NifH was done by using the 1HYQ Protein Data Bank (PDB) structure of MinD (Cordell and Lowe, 2001) and 1FP6 PDB structure of NifH (Jang et al., 2000). Similarly, for the superimposition of ArsA on NifH, the 1IHU PDB structure of ArsA (Zhou et al., 2001) was used. Structural models for CompA were generated using the 1HUX PDB structure (Locher et al., 2001).

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are described in Table 5.1. *E. coli* strains were normally grown at 37 °C in 2YT media

(Sambrook et al., 1992). Ampicillin was used to a final concentration of 50 µg/ml, wherever the selection was necessary.

Table 5.1. Bacterial strains and plasmids used in this study

Strain/plasmid	Relevant characteristics and description	Source/reference
Escherichia coli TOP10	F¯*mcr*A □ (*mrr-hsd*RMS-*mcr*BC) □ 80*lacZ*□M15 □*lac*X74 *rec*A1 *ara*D139 □(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen, CA
Escherichia coli JM109	*recA1 thi hsdR17supE44(lac-proAB) (*F' *traD36 proAB lacI*^{q} 1acZM15)	(obtained from Dr. Bhattacharjee and Dr. Rosen, Wayne State University, MI)
PCR 2.1 TOPO	*Amp', Kan',* (3.9kb), used for direct cloning of PCR products, *lacZá* fragment, *MCS, M13*	Invitrogen, CA
$pA_{H6}B$	Derivative of pALTER1 plasmid in which the 1749 bp *arsA* gene followed by six histidine codons and further followed by the 1287 bp *arsB* gene was cloned in its *HindIII-KpnI* site.	(obtained from Dr. Bhattacharjee and Dr. Rosen, Wayne State University, MI) (Li and Rosen, 2000)
pBG1757	Derivative of TOPO 2.1 in which an 876 bp DNA fragment corresponding to the linker+*arsA2*, followed by a 1287 bp *arsB* fragment was cloned directly after PCR	This work
pBG1791	Derivative of TOPO 2.1 in which a total of 3093 bp including the 876 bp *arsA2* DNA fragment (with linker) is fused 3' in frame to the first 876 bp of the *nifH* gene to encode a 584-residue NifH-ArsA2 chimeric protein. The 1287 bp *arsB* fragment follows the *arsA2* fragment.	This work
BG1754	Derived from *E. coli* JM109 strain by transformation using $pA_{H6}B$	This work
BG1757	Derived from *E. coli* JM109 strain by transformation using pBG1757	This work
BG1791	Derived from *E. coli* JM109 strain by transformation using pBG1791	This work

Overlap extension PCR. A hybrid NifH-ArsA2 protein was constructed by an overlap extension PCR method (Wurch et al., 1998) described as follows: the *nifH* gene (876 bp) was PCR amplified from *A. vinelandii* wild type cells using the oligonucleotides P1: 5' AAGCTTG CTATGCGTC AATGCGCCATCTAC 3' and P2: 5' GGCGACGCTACAGGCTGAGTG GAGACTTCTTCGGCGGTTTTG C 3'. For PCR amplifying the 2217 bp DNA fragment that encoded the ArsA linker followed by the ArsA2 and the ArsB (LA2B), the $pA_{H6}B$ plasmid was used as a template (kind contribution of Dr. Barry P. Rosen, Wayne State University, Detroit, MI). The $pA_{H6}B$

plasmid is a derivative of the pALTER1 plasmid, in which the *arsA* gene is followed by six histidine codons, further followed by the *arsB* gene (Li and Rosen, 2000).

The primers used for PCR amplifying the LA2B region were P3: 5' CGGCAAAACCGCCGAAGAAGTCTCCACTCAGCCTGTAGCG TCG 3' and P4: 5' AAGCTTTTACAATGTGACAG AGAGACGTAG. P2 was designed as a hybrid primer such that its initial 23 bp represented the start of the *arsA* linker sequence and the latter 20 bp represented the 3' *nifH* gene sequence (*nifH* stop codon omitted in primer design). P3 was also designed as a hybrid primer such that its initial 22 bp represented the 5' *nifH* end sequence and the latter 21 bp represented the 5' *arsA* linker sequence. Thus, P2 and P3 were designed so as to possess nucleotide extensions complementary to the end of the adjacent fragment of the chimeric gene, necessary to fuse the *nifH* and LA2B fragments together. In a second PCR step that utilized the external P1 and P4 primers and the products of the first PCR as templates, the entire *nifH*-LA2B fusion product was obtained. Due to the internal hybridization of the overlapping regions of the PCR products generated in the first PCR, a chimeric DNA fragment including the *nifH* and *arsA2B* was synthesized in the second PCR. For complete details of the overlap extension PCR reaction components and cycles, please refer the original article (Wurch et al., 1998).

Cloning and sequence verification. The chimeric DNA fragment obtained from the second PCR step was then cloned into the TOPO 2.1 plasmid vector and the construct was designated as pBG1791. The plasmid pBG1791 was sequenced for verification of the fusion between *nifH* and *arsA2* from the University of Chicago DNA Sequencing Facility

(Chicago, IL). Thus, in pBG1791, a total of 3093 bp including the 876 bp *arsA2* DNA fragment (with linker) is fused 3' in frame to the first 876 bp of the *nifH* gene to encode a 584-residue chimeric protein (292 residues of the NifH protein plus 292 residues of the ArsA2 protein (with linker)). The ArsB protein is transcribed as a single unit with the chimeric protein but is translated differentially. The recombinant NifH-ArsA2 protein is fused at its N-terminus with the first 26 amino acid residues of LacZ□, expressed from the vector. The pBG1791 plasmid was transformed into the *E. coli* JM109 cells and the resultant transformants were designated as BG1791. The 2217 bp *ars* LA2B fragment was also cloned separately into the TOPO 2.1 plasmid vector (pBG1757) after PCR amplification using primers P5: 5' AGATCTTCCACTCAGCCTGTAGCGTC GCCATCC 3' and P4 (3'*arsB*) and transformed into *E. coli* JM109 so that it could serve as a comparison for arsenite resistance measurements.

Genetic complementation. *E. coli* JM109, BG1754 (*arsAB*), BG1757 (*linker-arsA2B*) and BG1791 (*nifH-linker arsA2B*) cells were grown overnight at 37°C in LB media. The cells were then harvested and inoculated in triplicates into 5ml LB media tubes containing 0, 1, 3, 5 and 7 mM sodium arsenite and 0.1 mM IPTG and grown for 20 hours at 37°C with aeration. At 20 hours, the effect of the sodium arsenite on the growth of the cells was estimated by recording the cell density using O.D. $_{600}$. For detecting arsenite resistance on LB agar media, the *E. coli* JM109, BG1757, BG1791 and BG1754 strains were spotted after serial dilution, on plates containing 0, 2.5 and 5 mM sodium arsenite.

Growth in arsenite containing media. To observe and compare the growth rate of BG1754, BG1757 and BG1791, overnight cultures were grown, harvested and inoculated

into LB media tubes containing 5 mM sodium arsenite, so that the starting O.D. $_{600}$ was 0.04. Readings were taken at four-hour intervals, from 0 to 16 hours.

SDS-PAGE and Western Blot Analysis. A portion of each culture was centrifuged (0.6 ml) and the pellets were suspended in 0.1 ml of SDS sample buffer and boiled for 10 minutes. 15µl portions were analyzed by SDS-PAGE on 12% polyacrylamide gels and immunoblotted with anti-his antibody (directed against the 6X his tag at the C-terminus of the ArsA2 peptide).

Phylogenetic analysis. For the phylogenetic analyses of NifH and ArsA, related amino acid sequences were acquired from the GenBank database. The MEGA 3.1 software (Kumar et al., 2004) was utilized to generate the phylogenetic tree. The NifH amino acid sequences used in the phylogenetic analyses (GenBank protein accession numbers are in parentheses) were: *Anabaena variabilis* ATCC 29413 (YP_324741.1), *Azoarcus sp*. BH72 (AAG35586), *Azotobacter vinelandii* (AAA64709), *Bradyrhizobium japonicum* USDA 110 (NP_768409), *Burkholderia vietnamiensis* G4 (ZP_00420837), *Delftia tsuruhatensis* (AAS55953), *Desulfitobacterium hafniense* Y51 (YP_520504), *Erwinia carotovora* (CAG75856), *Frankia sp*. (CAA52161), *Geobacter sulfurreducens* (AAR36215), *Halorhodospira halophila* (BAD93282), *Heliobacterium chlorum* (BAD95753), *Herbaspirillum sp*. (BAD77945), *Klebsiella pneumoniae* (AAO85881), *Leptospirillum ferroxidans* (AAQ12257), *Magnetococcus sp.* (EAN30045), *Pelobacter propionicus* (ZP_00676010), *Rhizobium etli* (NP_659736.1), *Xanthobacter autotrophicus* (ZP_01197044) and the ArsA amino acid sequences used were: *Acidiphilium multivorum* (BAA24822.1), *Azoarcus sp.* EbN1 (YP_159868), *Bacillus sp.* CDB3 (AAD51849.1),

Chlorobium tepidum TLS (NP_662822), *Desulfitobacterium hafniense* DCB-2 (ZP_01372439), *Dictyostelium discoideum* (AAL96261), *Escherichia coli* (AAL96261) ArsA1, *Escherichia coli* (P52145) ArsA2, *Halobacterium sp*. NRC-1 (AAC82907), *Haloquadratum walsbyi* (YP_658327), *Idiomarina loihiensis* (YP_155089), *Klebsiella oxytoca* (AAF89640), *Lactobacillus plantarum* (CAG17843), *Leptospirillum ferroxidans* (AAY85169), *Methanococcus maripaludis* S2 (NP_987283), *Mus musculus* (AAB94772), *Psychrobacter arcticus* (YP_264406), *Rhodopirellula baltica* (CAD76332), *Shewanella sp*. ANA-3 (AAO31598), *Staphylococcus epidermidis* (YP_189973). The neighbor joining method was used for generation of the phylogenetic tree. Percentages from 500 replicate bootstrapping analyses were obtained for increased reliability.

Results and Discussion

ChlL: light-independent chlorophyll biosynthesis. The reduction of the C17-C18 double bond of the D-ring of protochlorophyllide (PChlide) to chlorophyllide (Chlide) is a major step in the chlorophyll biosynthesis pathway. The two pathways for this process involve catalysis by the enzymes NADPH : PChlide oxidoreductase (POR) and the light-independent PChlide reductase (DPOR). The light-independent PChlide reductase plays an important role in the ability of anoxygenic photosynthetic bacteria, cyanobacteria, nonvascular plants, ferns and gymnosperms to form chlorophyll in the dark (Fujita, 1996; Armstrong, 1998). Genetic and sequence analysis have suggested that the dark protochlorophyllide reductase (DPOR) consists of three protein subunits, ChlL, ChlB and ChlN as studied in cyanobacteria, or BchL, BchN and BchB as studied in bacteria.

Significant sequence similarity between the putative BchL/ChlL, BchN/ChlN and BchB/ChlB DPOR subunits has been found with the NifH, NifD and NifK subunits of nitrogenase respectively (Fujita, 1996; Fujita et al., 1993). Maximum similarity was found between NifH and ChlL structures, with an overall identity of 34% and a similarity of ~50% (Table 5.2).

Table 5.2. Percentage identity and percentage similarity (in parenthesis) of NifH with ChlL, MinD, ArsA1, ArsA2 and CompA

	NifH	ChlL	MinD	ArsA1	ArsA2	CompA
NifH	100% (100%)	34% (50%)	15.3% (36%)	17.1% (36%)	15% (32%)	11% (38.5%)
ChlL	34% (50%)	100% (100%)	17.7% (38.9%)	14.3% (32%)	15.6% (32%)	14.7% (34.4%)
MinD	15.3% (36%)	17.7% (38.9%)	100% (100%)	17.7% (31.4%)	18.1% (33%)	11.5% (40%)
ArsA1	17.1% (36%)	14.3% (32%)	17.7% (31.4%)	100% (100%)	25% (36.5%)	10% (36.6%)
ArsA2	15% (32%)	15.6% (32%)	18.1% (33%)	25% (36.5%)	100% (100%)	12.2% (37.4%)
CompA	11% (38.5%)	14.7% (34.4%)	11.5% (40%)	10% (36.6%)	12.2% (37.4%)	100% (100%)

Important features found conserved by comparison were the ATP-binding motif and the two Cys residues (C95 and C129) involved in coordinating the [4Fe-4S] cluster, in addition to two Asp residues postulated to have a role in ATP hydrolysis (Suzuki and Bauer, 1992; Huang and Liu, 1992; Burke et al., 1993), (Table 5.3), suggesting that the BchL/ChlL proteins might catalyze ATP-dependent transfer of electrons from a reductant such as ferredoxin, to a catalytic protein complex via the Fe:S center (Fujita and Bauer, 2000).

Table 5.3. Structural-functional features of NifH, ChlL, MinD, ArsA1, ArsA2 and CompA.

	NifH	ChlL	MinD	ArsA	CompA
localization	cytoplasmic	chloroplast	Peripheral membrane	Extrinsic membrane protein	cytoplasm
O2 sensitivity	Sensitive	Sensitive	Insensitive	Insensitive	Sensitive
Molecular weight	~64 kDa	~64 kDa (?)	~35 kDa (monomer)	~63 kDa	~54 kDa
Metal center	4Fe4S	4Fe4S (?)	–	SbIII/AsIII	4Fe4S
Folding pattern	3 layers: □/□/□ parallel or mixed beta -sheets of variable sizes	3 layers: □/□/□ parallel or mixed beta -sheets of variable sizes (?)	3 layers: □/□/□ parallel or mixed beta -sheets of variable sizes	3 layers: □/□/□ parallel or mixed beta - sheets of variable sizes	3 layers: □/□/□; mixed beta-sheet of 5 strands; strand 2 is antiparallel to the rest
Dimerization	Homodimer	Homodimer(?)	ATP dependent dimerization	Pseudodimer formed by connection of ArsA1 and ArsA2 domains through a short linker	Homodimer
Metal center ligands	C97, C132	C95, C129	–	H148 (A1), S420 (A2) C113 (A1), C422 (A2) C172 (A1), H453 (A2)	C127, C166
Nucleotide binding domain	10GKGGIGKSTTT 20	7GKGGIGKSTTS 17	9GKGGVGKTTSS 19	15GKGGIGKTSIS 25 (A1) 334GKGGVGKTTMA 344 (A2)	7GIDVG 11 100VIDIG 104
SwitchI - SwitchII	39DPKADST45 (Switch I) 125DVLGDVV 131 (Switch II)	37DPKHDST 43 (Switch I) 122DVLGDVV 128 (Switch II)	43DFDIGLRNL 51 (Switch I) 120DSPAG 124 (Switch II)	45DPASN 49 (A1 Switch I) 365DPAAH 369 (A2 Switch I) 142DTAPGHT 149 (A1 Switch II) 447DTAPGHT 453 (A2 Switch II)	104GGQD 107 127CAAGTG 132 (Switch II)

Based on the sequence similarity between DPOR and nitrogenase, Fujita and Bauer suggested that the BchL protein existed in solution as a dimer (Fujita and Bauer, 2000). The reconstitution of the DPOR subunits showed that it consists of two separable components, the BchL protein and a BchN-BchB-protein complex, similar to that observed for nitrogenase, which has separable NifH and NifD-NifK components (Fujita and Bauer, 2000). Recently, the functional substitution for the *chlL* gene in

Chlamydomonas reinhardtii was achieved by its replacement with the *nifH* gene of *Klebsiella pneumoniae* (Cheng et al., 2005). The ability of NifH to functionally substitute for ChlL enhanced the view that the two proteins bear high structural equivalence to each other. Also, there are ongoing attempts towards substituting the *nifH* gene with the *chlL* gene and determining its sufficiency in the nitrogenase complex (Gavini *et al*., unpublished data). The presence of a [4Fe-4S] cluster in ChlL is an assumption based on its similarity to NifH. The accessory gene products required for the mobilization of sulfur and iron for Fe-S cluster formation and for the activity and stabilization of NifH include NifS, NifU and NifM (Gavini and Pulakat, 2002; Lill and Muhlenhoff, 2005). An inspection of the *Chlamydomonas* genome database for the equivalents of these accessory proteins did reveal their matches, leading to further evidence of a Fe-S cluster in ChlL (Cheng et al., 2005).

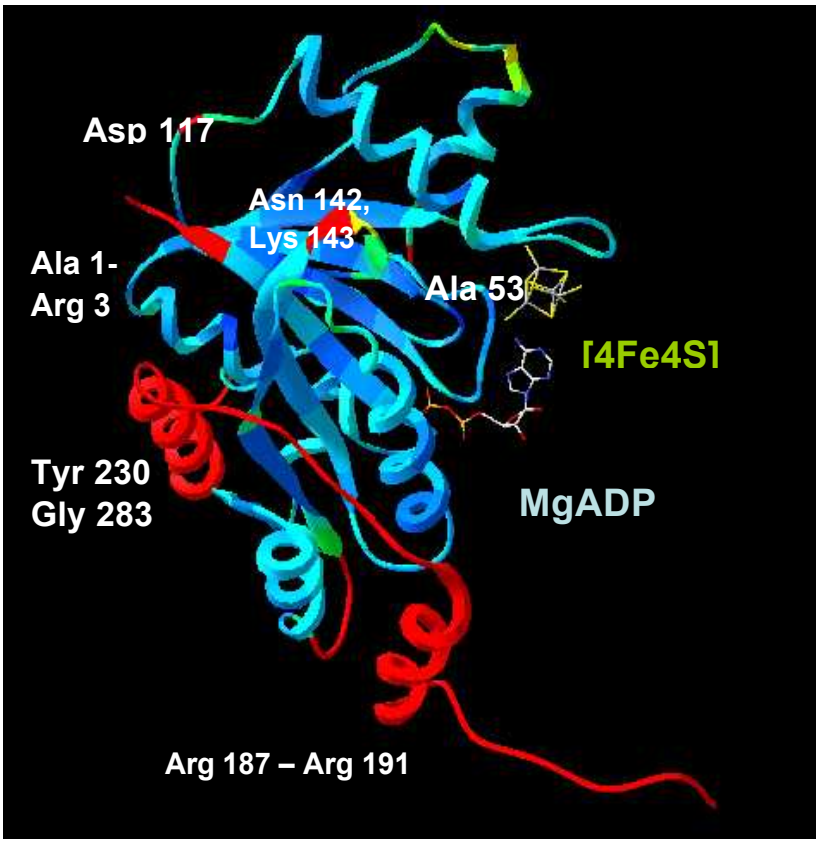
Inspite of many genetic and biochemical investigations of DPOR, the lack of a crystallized ChlL protein has deterred a detailed structural insight into the protein. As a step further to elucidate the structural aspects of the ChlL protein, we superimposed the protein sequence derived ChlL onto the NifH template (PDB ID: 1NIP), using the Swiss PDB (Deep View) protein modeling software (Fig. 5.3). The resultant ChlL protein model generated showed marked resemblance to the NifH overall architecture and folds. As shown in Fig. 5.4, upon superimposition of a ChlL monomer on the NifH monomer, there was a wide overall structural alignment, however, the following residues in NifH appeared non-aligned: A1-R3, A53, D117, N142, K143, R187-R191, Y230-G283. Particularly, the ‘tail region’ of NifH, comprised of residues 230-283 (Fig. 5.4), differed

extensively, since no such region of ChlL could be structurally aligned to it. Such absence of a tail region in ChlL could explain their phylogenetic divergence and possibly account for their functional differences.

The model shows □-helices colored in maroon and â-strands colored in light yellow. The model depicted here is only of a monomeric unit of the ChlL protein.

Fig. 5.3. A 3-D model of the ChlL protein generated by using NifH as a template

Besides ChlL, CompA is another oxygen sensitive protein that has structural similarity to NifH and is a homodimer containing a [4Fe-4S] cluster (Table 5.3). CompA shares ~38.5% protein sequence similarity with NifH (Table 5.2).



The superimposition of ChlL on NifH indicated only a few regions of dissimilarity (shown in red). A root mean square deviation type of coloration of the NifH protein (in comparison to ChlL) shows a particularly dissimilar 'tail' region (red stretch) spanning residues 230-289 of NifH. Each dissimilar region is indicated by the residues comprised therein. The [4Fe4S] cluster and the bound MgADP in NifH are indicated.

Fig. 5.4. NifH monomer structure colored according to differences with the ChlL model.

CompA: Glutamate degradation. The hydroxyglutarate pathway, which involves the *syn*-elimination of water from (*R*)-2-hydroxyglutaryl-CoA is key to glutamate degradation in *Acidaminococcus fermentas* (Buckel, 1980; Muller and Buckel, 1995). This important step in the process of glutamate degradation is catalyzed by the enzyme 2-hydroxyglutaryl-CoA dehydratase. This enzyme is the product of the *hgdCAB* genes and consists of the protein components, component A (CompA) and component B (CompD).

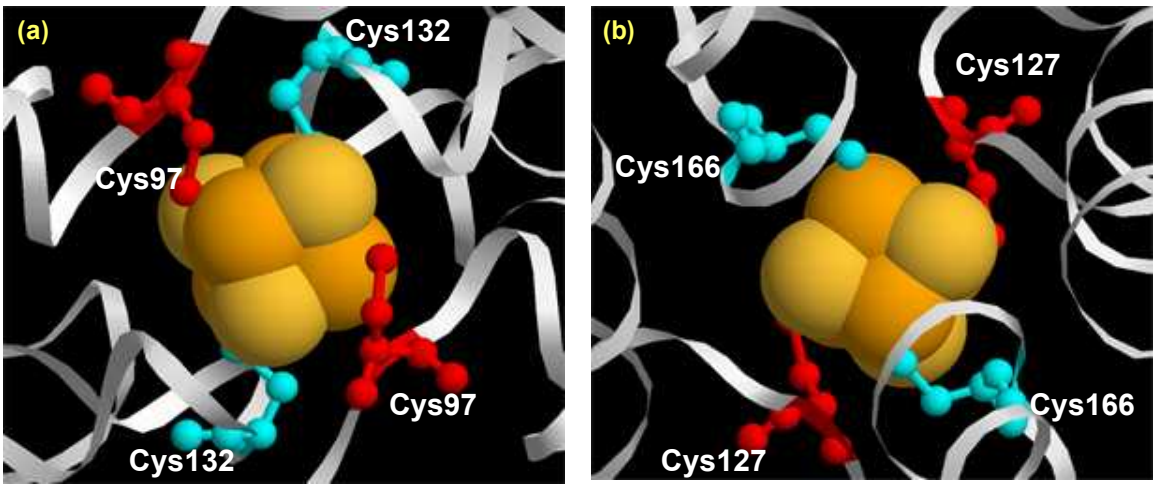
CompA is an extremely oxygen-sensitive activator or initiator whereas CompD is moderately oxygen-sensitive and is the actual dehydratase. For activity of the enzyme, both protein components, Mg^{2+}, ATP and a strong reducing agent are required. ATP is bound and hydrolyzed by CompA whilst the substrate binds to CompD (Hans et al., 1999). The X-ray crystallized structure of recombinant CompA protein from *A. fermentas* was determined at 3 Å resolution by Locher *et al*. in 2001 (Locher et al., 2001). The CompA protein resembles an 'L' shape, with its two arms being roughly similar in length. It has eight □-helices and ten â-strands that form two flat sheets (Locher et al., 2001). These two sheets meet at an angle of 60° and the second sheet positions at the corner of the 'L' and therefore is at the center of the protein. Helices 2 and 8, which are almost perpendicular to each other, fill up the large gap between the two â-sheets (Locher et al., 2001). The ADP is bound between the two edges of the closely situated â-sheets. The â-phosphate of the ADP is positioned ~19 Å from the [4Fe-4S] cluster that lies at the dimeric interface of CompA (Locher et al., 2001). The nucleotide binding pocket in CompA is formed of residues in strands 1 and 6, containing the Gly7-X-Asp-X-Gly11 and the Val100-X-Asp-X-Gly104 motif (Table 5.3), as found in sugar kinases (Hurley, 1996).

Though NifH and CompA are from unrelated phylogenetic families, several structural resemblances exist between them. Like NifH, CompA is a homodimeric enzyme (54 kDa), with one ADP bound to each subunit and a single [4Fe-4S] cluster at its dimeric interface. The [4Fe-4S] cluster in both, CompA and NifH, are solvent exposed (~20 $Å^2$) instead of buried, as observed in other [4Fe-4S] cluster containing

proteins (Locher et al., 2001; Strop et al., 2001). Excluding the cluster, the interface between CompA monomers accounts for only ~800 $Å^2$ buried area (Locher et al., 2001). From predictions based on protein sequence comparisons and subsequent structural analysis of the X-ray crystallized CompA, the [4Fe-4S] cluster was found to be coordinated by two cysteine residues, C127 and C166, provided by each monomer (Locher et al., 2001). The coordination of the Cys ligands with the [4Fe-4S] cluster in NifH and CompA is shown in Fig. 5.5a and 5.5b respectively. Two helices, each from one subunit, pointing with their N-termini towards the [4Fe-4S] cluster and forming a helix-cluster-helix angle of 105°, is a remarkable feature of CompA Kim et al., 2004). Such similar architecture has also been noted in NifH, revealing a helix-cluster-helix angle of 150° (Howard and Rees, 1994). Further, as observed in the NifH complex with MoFe protein in the presence of ADP-AlF_4^-, the CompA also probably opens to an angle of 180° upon ATP binding. This 'open' structure then resembles an archer shooting arrows, wherein the helix-cluster-helix with two bound ADP could be superimposed on the 'string of the archer's bow', hence leading to the term 'archerase' for CompA (Buckel, 2003).

The Switch II region in NifH includes the Cys132 residue that coordinates the [4Fe-4S] cluster and consequently links the nucleotide and the cluster in NifH (Schindelin et al., 1997). The Switch II region interacts with the □-phosphate of the nucleotide and upon hydrolysis changes its conformation to the nucleoside diphosphate state. Thus, in NifH, the regulation of electron transfer by protein conformational change is directed by the nucleotide hydrolysis. In CompA, a structural analogue equivalent to

the NifH Switch II region has been mapped to the residues 104 to 107 and 127 to 132, positioned between the ADP and [4Fe-4S] cluster; these segments connect the nucleotide-binding site to the cluster and have been suggested to be involved in coupling ATP hydrolysis to the changes in the cluster-binding region (Locher et al., 2001).



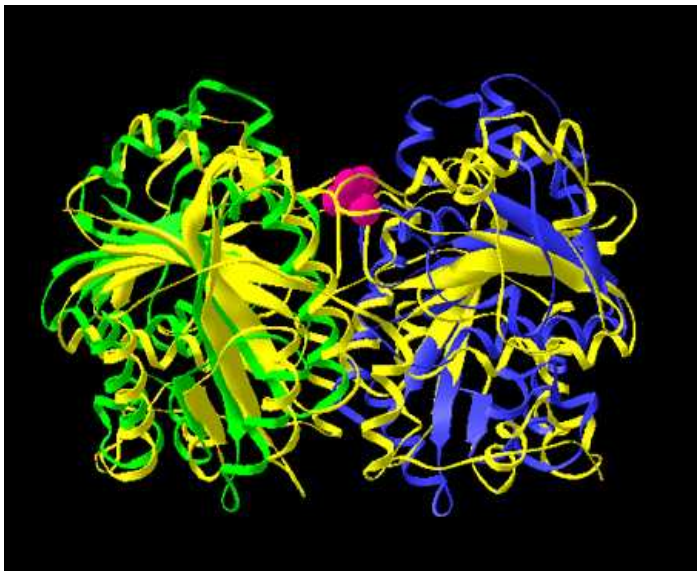
(a) In NifH, Cys 97 (ball and stick- red) and Cys 132 (ball and stick- cyan) serve as ligands for the [4Fe4S] cluster (spacefill-CPK) (b) In CompA, Cys 127 (ball and stick- red) and Cys 166 (ball and stick- cyan) serve as ligands for the [4Fe4S] cluster (spacefill-CPK).

Fig. 5.5. Similar view of [4Fe4S] cluster bound by Cys ligands in NifH and CompA

Within this group of NifH similar proteins (Table 5.3), MinD and ArsA are marked by the absence of the [4Fe-4S] cluster. The ArsA protein could be considered to be a dimer of two NifH like subunits ArsA1 and ArsA2; however for MinD, only a monomeric form has been available for study by researchers, although evidence for ATP dependent dimerization of MinD has been found (Hu et al., 2003).

MinD: spatial regulator of cell division. The peripheral membrane ATPase MinD, along with MinC and MinE, is a component of the Min system responsible for correct placement of the division site in *E. coli* cells (Rothfield et al., 2001). MinD helps in blocking unnecessary septation events at the poles by rapidly migrating from one cell pole to the other. The binding of MinE to MinD induces hydrolysis of ATP and the release of MinD into the cytoplasm (Hu et al., 2002). The induction of MinE by the ATP binding cycle results in the rapid movement of MinD from one cell pole to the opposite cell pole, forming alternating broad polar zones (Raskin and de Boer et al., 1999). It was recently reported that MinC and MinD, which were previously thought to be diffusely associated with the membrane at the poles, were in reality organized into extended spirals that coiled around the cell (Shih et al., 2003). Although it has been shown that MinD dimerizes in the presence of ATP, only its monomeric form with ADP, AMPPCP or without a nucleotide has been crystallized till now (Hayashi et al., 2001; Cordell and Lowe, 2001; Sakai et al., 2001). These crystallized MinD were obtained from three different archaeal species, *Pyrococcus horikoshii* OT3, *Pyrococcus furiosus* and *Archaeoglobus fulgidus* based on the relative ease of their crystallization. The MinD molecule from *P. horikoshii* was found to consist of a â-sheet comprised of 7 parallel and 1 antiparallel strands and 11 peripheral á-helices (Sakai et al., 2001). Though NifH and MinD share only ~16% identical residues (Table 5.2), the overall folding topology of the two proteins are clearly related to each other (Sakai et al., 2001). Since the dimeric form of crystallized MinD was unavailable, Lutkenhaus and Sundaramoorthy used the NifH structure as a template for modeling a dimeric MinD (Lutkenhaus and Sundaramoorthy, 2003). A similar image

was generated for the purpose of this review using the crystallized MinD (PDB ID: 1HYQ) and NifH (PDB ID: 1NIP) structures as shown in Fig. 5.6. The superimposition of two MinD monomers (green and blue) on a NifH dimer template (yellow) revealed that other than certain loops at the surface of the two proteins, an additional loop and an á-helix present near the C-terminus in NifH, the superimposition yielded a good fit, similar to the observation of Lutkenhaus and Sundaramoorthy (Lutkenhaus and Sundaramoorthy, 2003). Likewise, Hayashi *et al*. found that NifH and MinD exhibited



The MinD monomers (blue and green) have been superimposed on a NifH dimer (yellow). The [4Fe4S] cluster of NifH is shown as a spacefill body (magenta).

Fig. 5.6. Comparisons between NifH and MinD.

similar architectural regions in their ATPase and GTPase domains (â7, â6, â1, â5, â2, â4, â3, the P-loop, á1, á7, á8) yielding 1.2 Å r.m.s.d. value for 81 C á atoms (Hayashi et al., 2001).

Both NifH and MinD contain the deviant Walker motif A [X-K-G-G-X-X-K-(T/S)] (Koonin, 1993). It was however observed that the C97 and C132 ligands of the [4Fe-4S] cluster of NifH were not conserved in *P. furiosus* MinD. Upon protein sequence comparison of NifH and MinD from *A. fulgidus*, a distinct presence of the Fer4_NifH conserved domain, typically found in [4Fe-4S] iron-sulfur cluster binding proteins and NifH/frxC protein families, was detected in MinD (Marchler-Bauer, 2005), spanning residues 133-240 of MinD that correspond to residues 144-251 of NifH (Fig. 5.7 and 5.8).

ArsA: oxyanion translocating ATPase subunit. The ArsA protein that is the catalytic subunit of the As(III)/Sb(III)-translocating ArsAB complex, has long been speculated to be a distant relative of NifH (Gatti et al., 2000). The ArsB subunit serves as a membrane

```
1HYQ 133 AQELLLVVNPEISS------ITDGLKTKIVAERLGTKVLGVVVN-RITTLGIEMAKNEIE 193
1NIP 144 AQEIYIVCSGEMMAMYAANNISKGIVKY--ANSGSVRLGGLICNSRNTDREDELIIALAN 204

1HYQ 194 AILEAKVIGLIPEDPEVRRAAAYGKPVVLRSPNSPAARAIVELANYI 240
1NIP 205 KL-GTQMIHFVPRDNVVQRAEIRRMTVIEYDPKAKQADEYRALARKV 251
```

This Fer4_NifH domain was detected in residues 132-232 of MinD which correspond to residues 144 to 251 of NifH. Here 1HYQ indicates the PDB ID of MinD protein from *A. fulgidus* and 1NIP indicates the PDB ID of NifH from *A. vinelandii*. The alignment indicates unaligned residues in gray, conserved residues in red and deviant residues in blue

Fig. 5.7. The conserved domain found in 4Fe4S iron-sulfur cluster binding proteins and NifH/frxC protein families.

anchor for ArsA and forms the anion-translocating sector of the ArsAB pump (Dey et al., 1994). The crystallized structure of the ArsA protein (Zhou et al., 2000) has revealed many structural details that establish its similarity to the NifH protein. ArsA is a 583-

residue polypeptide that has homologous N-terminal (A1) and C-terminal (A2) halves, "indicating an evolutionary gene duplication and fusion" (Chen et al., 1986).

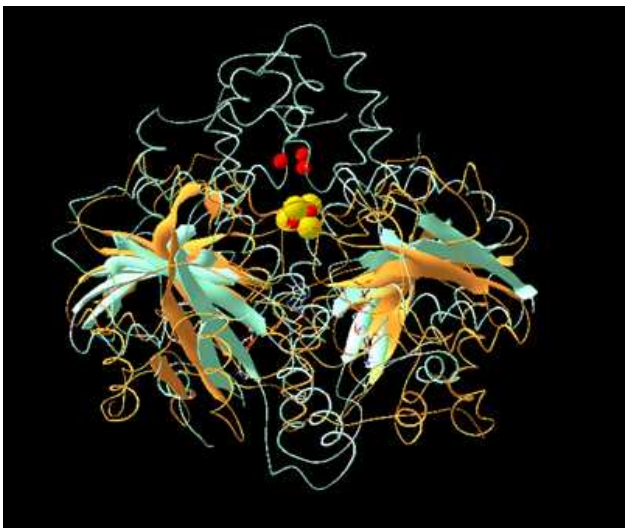
Fig. 5.8. Structural representation of the Fer4_NifH domain. The Fer4_NifH domain (residues 132-232) is depicted in cyan color in a 3-D representation of the MinD protein (PDB ID: 1HYQ). Remaining portion of the protein is colored according to presence of secondary structure (□-helices-magenta, â-strands-orange)

This 63kDa ATPase contains two consensus nucleotide-binding motifs, one each in the A1 and A2 halves. The A1 and A2 halves are held together by a short linker peptide (Li and Rosen, 2000). ArsA, like NifH exists in a more open conformation in the presence of ADP and contains two NBSs. The two NBSs are each filled with MgADP and lie at the interface between A1 and A2. In each NBS, Mg^{2+} is octahedrally coordinated by the □-phosphate of ADP, several water molecules and a threonine hydroxyl. When compared to D39 and D129 residues of the NifH protein that play a role in the MgATP hydrolysis, D45 from the ArsA A1 NBS directly coordinates with Mg^{2+}

whereas D364 of the ArsA A2 NBS coordinates indirectly to it via a water molecule (Zhou et al., 2000; Lanzilotta et al., 1995; Lanzilotta et al., 1997). ArsA is a homodimer in its catalytically active form (Kaur and Rosen, 1993) and dimerization is favored by the formation of a three-coordinate complex among three specific cysteine thiolates (Cys-113, Cys-172, and Cys-442) and the effector, Sb(III) or As(III) (Ching et al., 1991; Bhattacharjee et al., 1995). Each Sb(III) is coordinated by one A1 residue and one A2 residue. These residues include His148 (A1) and Ser420 (A2); Cys113 (A1) and Cys422 (A2); Cys172 (A1) and His453 (A2) (Zhou et al., 2000). The oxygen insensitivity of ArsA in comparison to NifH could most reasonably be assigned to the difference in the metal center of both these proteins

ArsA is twice the size of NifH, yet since it consists of two similar domains connected by a short linker, each ArsA monomer could actually be considered to be a pseudodimer, as regarded herein. Information obtained on ArsA (PDB ID: 1IHU) and NifH (PDB ID: 1G5P) protein families from the SCOP database (Murzin et al., 1995) and their comparison based on structural modeling made it evident that the core â-sheet pattern in both proteins, made up of at least 7 parallel â-strands, could fit with <2.5 Å root square deviation upon superimposition (Fig. 5.9). It was also observed that the Cys97 ligands for the [4Fe-4S] cluster from the two NifH monomers could be aligned to the Cys113 and Cys422 residues of ArsA A1 and ArsA A2 respectively (Fig. 5.10). Moreover, the Cys132 ligands for the [4Fe-4S] cluster of NifH corresponded to the His 148 (A1) and His453 (A2) of the antimony site. The superimposition of the ArsA A1 or A2 with a NifH monomer revealed a clear coincidence of the metalloid center of ArsA

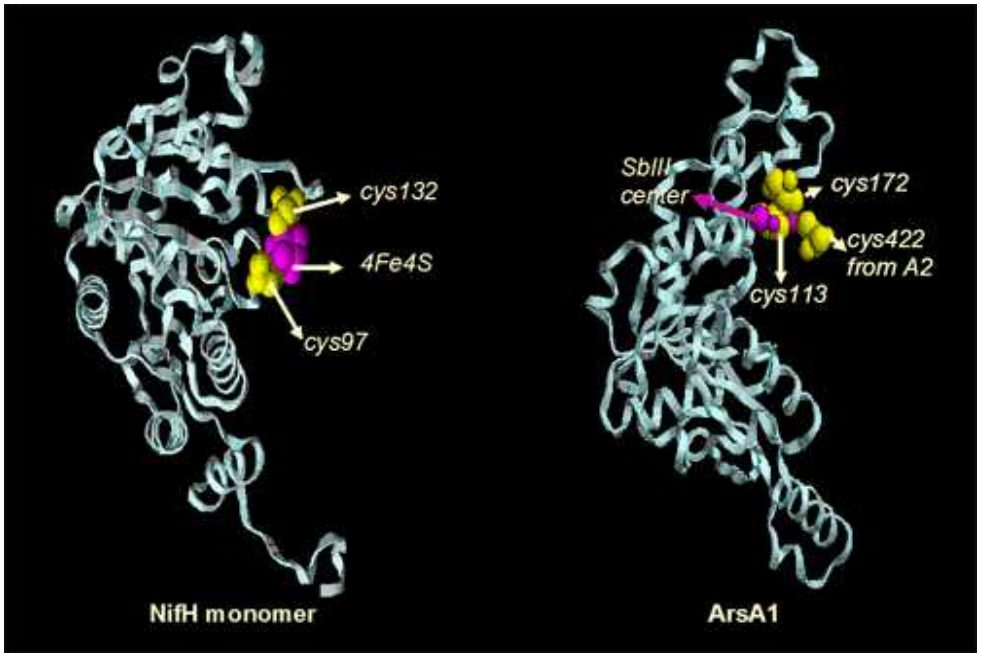
and the [4Fe-4S] cluster of NifH. Furthermore, the $D^{142/447}TAPTGH^{148/453}$ signature sequence of ArsA A1 and A2 (Zhou et al., 2000) has an exact counterpart in NifH and is believed to correspond to the Switch II region of G-proteins (Jang et al., 2000).



Superimposition of ArsA (cyan) on NifH (yellow) shows near coincidence of their metal clusters and indicates a good fit of the â-strand structures. [4Fe4S] cluster of NifH shown spacefilled in yellow and Sb(III)/As(III) metal center shown spacefilled in red

Fig. 5.9. Structural comparisons between ArsA and NifH.

The comparison of the residues of ArsA A1 and A2 corresponding to the region of NifH involved in complex formation with MoFe protein yielded a similar patch of acidic residues. E112 of NifH and K400 of the MoFe protein which specifically crosslink during this process are found within a pocket of other close acidic residues and basic residues respectively (Schmid et al., 2002). E112 in NifH is surrounded by multiple



Both protein structures are represented as cyan ribbons and the metal centers are shown in magenta (spacefill). Cys 97 and Cys 132 for NifH and Cys 113 and Cys 172 for ArsA (A2) are indicated by arrows. Also Cys 422 implicated as a ligand from ArsA (A1) is indicated.

Fig. 5.10. A comparison of the metal centers and ligands of NifH and ArsA in a similar view.

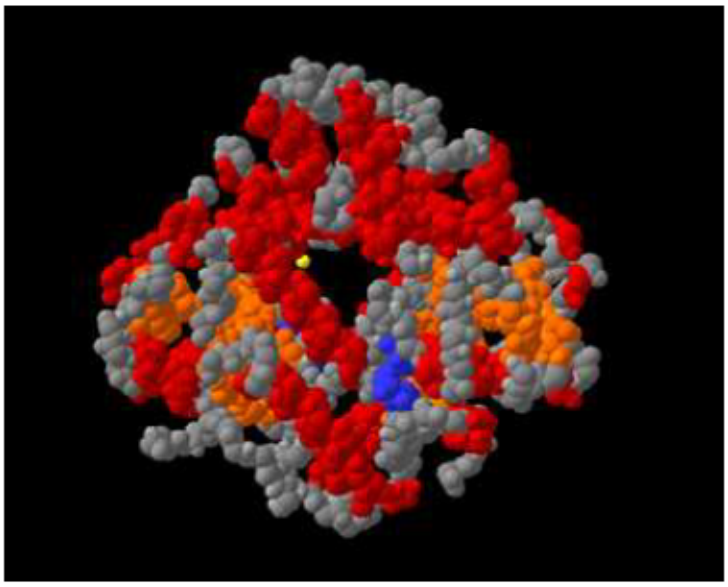
carboxylic acids, including E68, 71, 73, 110, 111, 112, 116 and D69 (Schmid et al., 2002). Although E112 is highly specific for the crosslinking, it is not a conserved residue among the NifH protein sequences. However, the acidic patch surrounding it is well conserved. The ArsA A2 region comprising of residues from L398 to G439, when compared with the region G94 to F135 of NifH, revealed eight glutamate residues, two of which had exact counterparts in NifH and E112 of NifH corresponded to D417 of ArsA A2 (Fig. 5.11).

```
                    100         110         120         130
NifH      94  GVGCAGRGVITAINFLEEEGAY DDLDFVFYDVLGDVVCGGF 135
ArsA2    398  L TKGK LD AGKRLLEEDLRSPCT  IAVFQAFSRVIR AG 439
```

Protein sequence alignment of NifH and Ars2 in the region corresponding to important residues in NifH required for complex formation with the MoFe protein (Gly94 to Phe135 in NifH and Leu398 to Gly439 in ArsA) is shown. An acidic patch can be observed in ArsA, indicated by the presence of eight glutamate residues (highlighted in black). Glu112 of NifH corresponds to Asp 417 of ArsA2. Asp417 could be also considered as Asp99 when ArsA2 is taken as a single unit.

Fig. 5.11. Conservation of putative docking residues.

The buried interface between the A1 and A2 domains of ArsA was found to be relatively small (<10%), compared with the total surface of the enzyme, rendering ArsA as a hollow protein with a large central cavity (Fig. 5.12).



A spacefilled structure of the ArsA protein is shown. □-helices are in red, â-strands are in orange, coils in gray and seen in yellow is part of the metal center

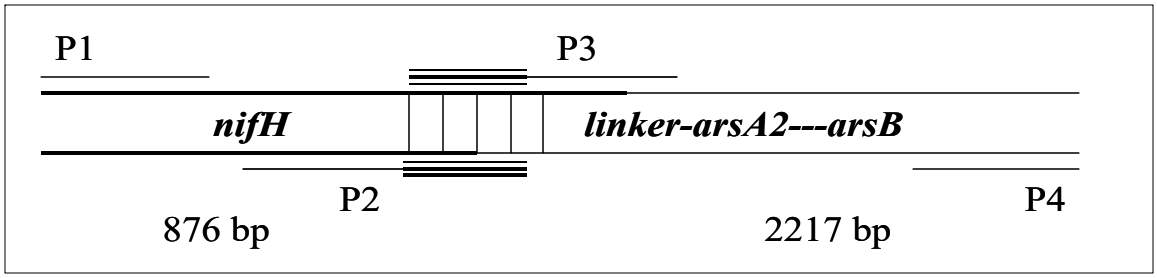
Fig. 5.12. The ArsA is a hollow protein with a central cavity.

The hinge region between the ArsA A1 and A2 polypeptides is formed by a 25-residue linker peptide in the region of 289-313 (Li and Rosen, 2000). The structural comparison of the linker region of ArsA by superimposition on NifH revealed the absence of such a stretch in NifH.

Functional substitution of the ArsA protein by the NifH-ArsA2 chimeric protein. Our hypothesis that a chimeric NifH-ArsA2 protein would be capable of conferring arsenite resistance to arsenite sensitive wild type *E. coli* cells, emerged from the observation of these striking structural and mechanistic similarities between NifH and ArsA. In order to test whether NifH could substitute for the ArsA1 half of the ArsA protein, we constructed a hybrid NifH-ArsA2 protein by an overlap extension PCR method (Wurch et al., 1998), as described in the *Materials and Methods* section. As shown in Fig. 5.13, in pBG1791 (plasmid containing *nifH-linker arsA2B*), a total of 3093 bp including the 876 bp *arsA2* DNA fragment (with linker) is fused 3' in frame to the first 876 bp of the *nifH* gene to encode a 584-residue chimeric protein (292 residues of the NifH protein plus 292 residues of the ArsA2 protein (with linker)). The ArsB protein is transcribed as a single unit with the chimeric protein but is translated differentially

The recombinant NifH-ArsA2 protein is fused at its N-terminus with the first 26 amino acid residues of LacZ□, expressed from the vector. The pBG1791 plasmid was transformed into the *E. coli* JM109 cells and the resultant transformants were designated as BG1791. The 2217 bp *ars* LA2B fragment was also cloned separately into the TOPO 2.1 plasmid vector (pBG1757) and transformed into *E. coli* JM109 so that it could serve as a comparison for arsenite resistance measurements. The pBG1757

transformed *E. coli* JM109 strain was designated as BG1757. The JM109 *E. coli* cells containing the $pA_{H6}B$ plasmid were designated as BG1754.



The P1 and P2 primers were used to amplify 876 bp of *nifH* and the P3 and P4 primers were used to amplify the region comprising *linker-arsA2-arsB*. These PCR products were used as templates for a second PCR step in which the entire *nifH-linker-arsA2-arsB* fragment was obtained after amplification with primers P1 and P4. The overlapping sequences of primers P2 and P3 are shown as dark double lines and the internal hybridization occurring in that region during the second PCR is indicated with vertical lines between *nifH* and the *linke*r of *arsA*.

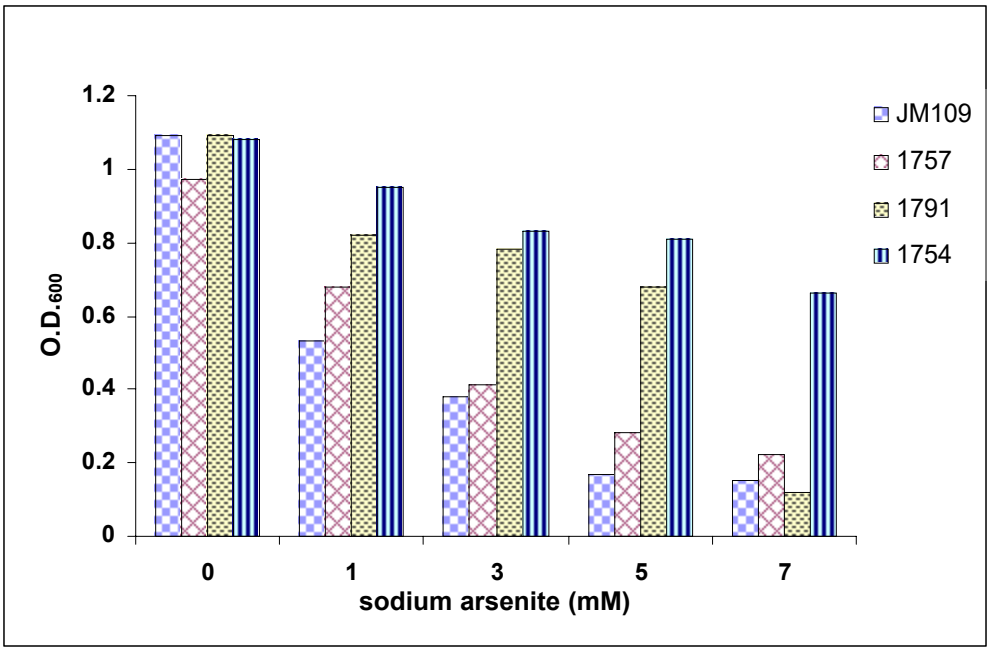
Fig. 5.13. Schematic of the *nifH-arsA2B* construct

We tested and compared the extent of arsenite resistance between the *E. coli* BG1757 (LA2B), BG1791 (*nifH*-LA2B chimera) and BG1754 (*arsAB*) strains. For observing the extent of arsenite resistance in liquid culture, the cells were first grown overnight at 37°C in LB media. The cells were then harvested and inoculated in triplicates into 5ml LB media tubes containing 0, 1, 3, 5 and 7 mM sodium arsenite and 0.1 mM IPTG and grown for 20 hours at 37°C with aeration. At 20 hours, the effect of the sodium arsenite on the growth of the cells was estimated by recording the cell density using $O.D._{600}$. As seen in Fig. 5.14, BG1791 was able to show resistance to arsenite upto a concentration of 5mM sodium arsenite, whereas, in comparison, BG1757 showed much

less growth, starting from 3mM sodium arsenite. At 1, 3 and 5 mM sodium arsenite concentrations, BG1791 showed the same pattern and extent of growth as the BG1754 strain containing the complete ArsA protein. In previous studies of arsenite resistance, it was usually found that cells expressing the wild type *arsAB* genes could tolerate upto 5-7 mM sodium arsenite (Kaur and Rosen, 1993; Li et al., 1996) and our results are comparable to them. Similar results were found on LB agar media, when the *E. coli* JM109, BG1757, BG1791 and BG1754 strains were spotted after serial dilution, on plates containing 0, 2.5 and 5 mM sodium arsenite (Fig. 5.15). To observe and compare the growth rate of BG1757, BG1791 and BG1754, overnight cultures were grown, harvested and inoculated into LB media tubes containing 5 mM sodium arsenite, so that the starting O.D. $_{600}$ was 0.04. Readings were taken at four-hour intervals, from 0 to 16 hours. Fig. 5.16 represents the growth curve pattern obtained from this experiment. It was seen that BG1791 had a large lag phase (~4 hours) but its ability to tolerate arsenite increased gradually and it showed comparable growth to BG1754 at 16 hours. All the above data were obtained from triplicates in each experiment and confirmed by performing the experiments 3 times each.

To determine whether a protein of the size of the recombinant NifH-ArsA2 (~64 kDa) protein was being produced in BG1791, we performed SDS-PAGE and western blot analysis. For the western blot, an anti-his antibody was used against the 6X his tag at the C-terminus of the ArsA2 peptide. As seen in Fig. 5.17, a protein band corresponding to the expected size of the NifH-ArsA2 (~64 kDa) was found to be expressed from the BG1791 cells (Fig. 5.17). Polypeptides corresponding to the ~47.6 kDa ArsA2 and ~63

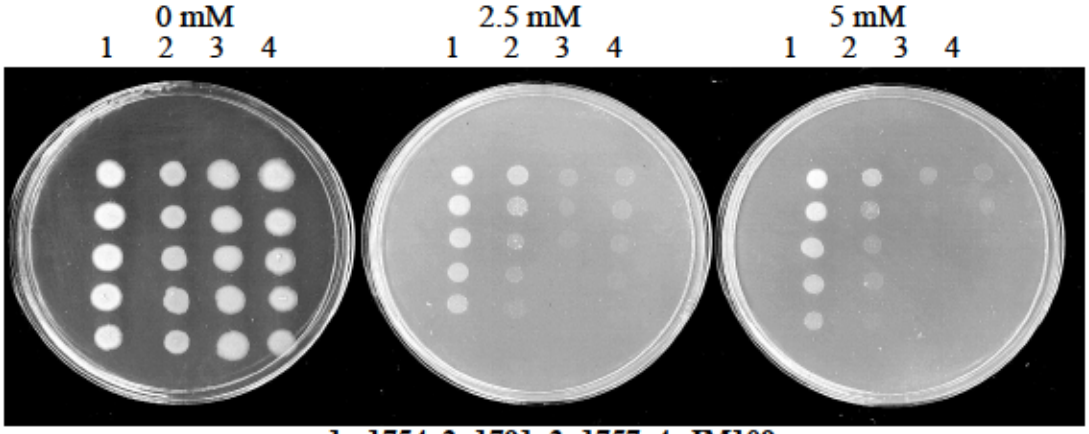
kDa ArsA were found to be expressed from the BG1757 and BG1754 cells respectively (Fig. 5.17).



Overnight cultures of *E. coli* strains JM109, BG1754 (*arsA*), BG1791 (*nifH-LA2B*) and BG1757 *(arsA2)* were diluted 100-fold into fresh LB medium containing varying concentrations of sodium arsenite and 0.1 mM IPTG. O. D. $_{600nm}$ was measured after 20 hours of growth at 37°C.

Fig. 5.14. Resistance to arsenite in cells expressing wild type (*arsA)*, chimeric (*nifH-LA2B)* and partial (*arsA2) arsA* genes.

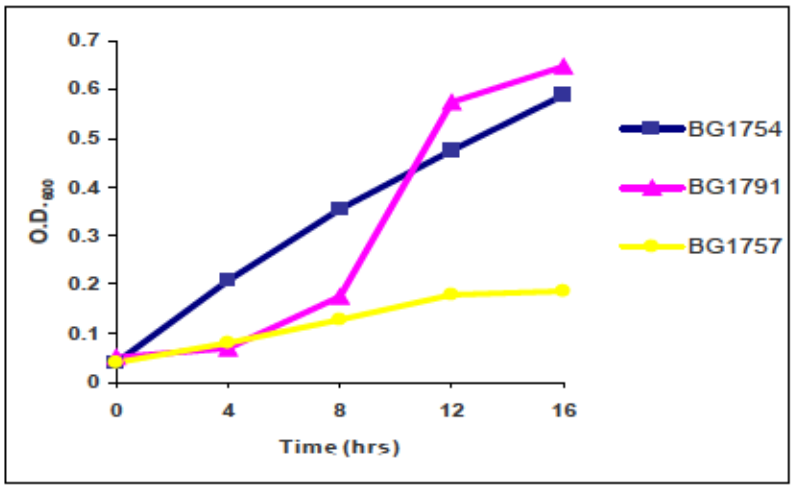
We further conducted phylogenetic analyses of the NifH and ArsA proteins to assess their evolutionary inter-relationships. NifH and ArsA related amino acid sequences were acquired from the GenBank database (refer *Materials and Methods* section for accession numbers). The MEGA 3.1 software (Kumar et al., 2004) was utilized to generate the phylogenetic tree.



1= 1754, 2=1791, 3=1757, 4=JM109

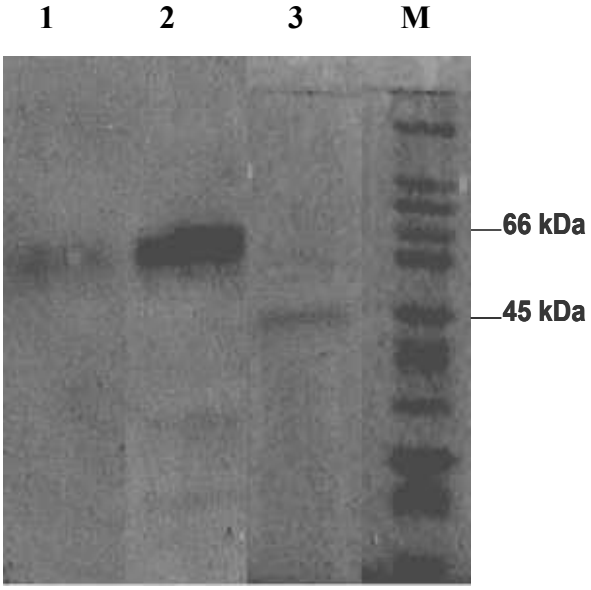
Resistance to arsenite in cells expressing wild type (*arsA)*, chimeric (*nifH-LA2B)* and partial (*arsA2) arsA* genes is shown. *E. coli* JM109, BG1754, BG1791 and BG1757 were grown till they reached an O. D. $_{600nm}$ of 1.2. The cells were serially diluted (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}) and 2□l of the undiluted and diluted cells were spotted on an LB agar plate supplemented with 50□g/ml ampicillin and 0.1 mM IPTG.

Fig. 5.15. Dilution spotting on LB agar media.



Overnight cultures were diluted 100-fold into fresh LB media containing 5mM concentration of sodium arsenite and 0.1 mM IPTG and grown at at 37°C with aeration. The cell densities of BG1754 (*arsA*), BG1791 *(nifH-LA2B)* and BG1757 (*arsA2)* were recorded at O. D. $_{600nm}$ for 16 hours at four hour intervals.

Fig. 5.16. Growth rate characteristics of *E. coli* cells expressing wild type (*arsA)*, chimeric (*nifH-LA2B)* and partial (*arsA2) arsA* genes.



Lanes: 1, NifH-ArsA2 chimera, 2, ArsA, 3, ArsA2. Arrows indicate the migration positions of the ~47.6 kDa ArsA2 and ~64 kDa NifH-ArsA2 and ArsA polypeptides, as determined from the M-4038 (Sigma, Saint Louis, MO) standard molecular weight marker (M). An anti-his antibody directed against the 6X his tag following the ArsA2 peptide was used in all three lanes.

Fig. 5. 17. Western blot analysis of expressed gene products from BG1757 (ArsA2), BG1754 (ArsA) and BG1791 (NifH-ArsA2).

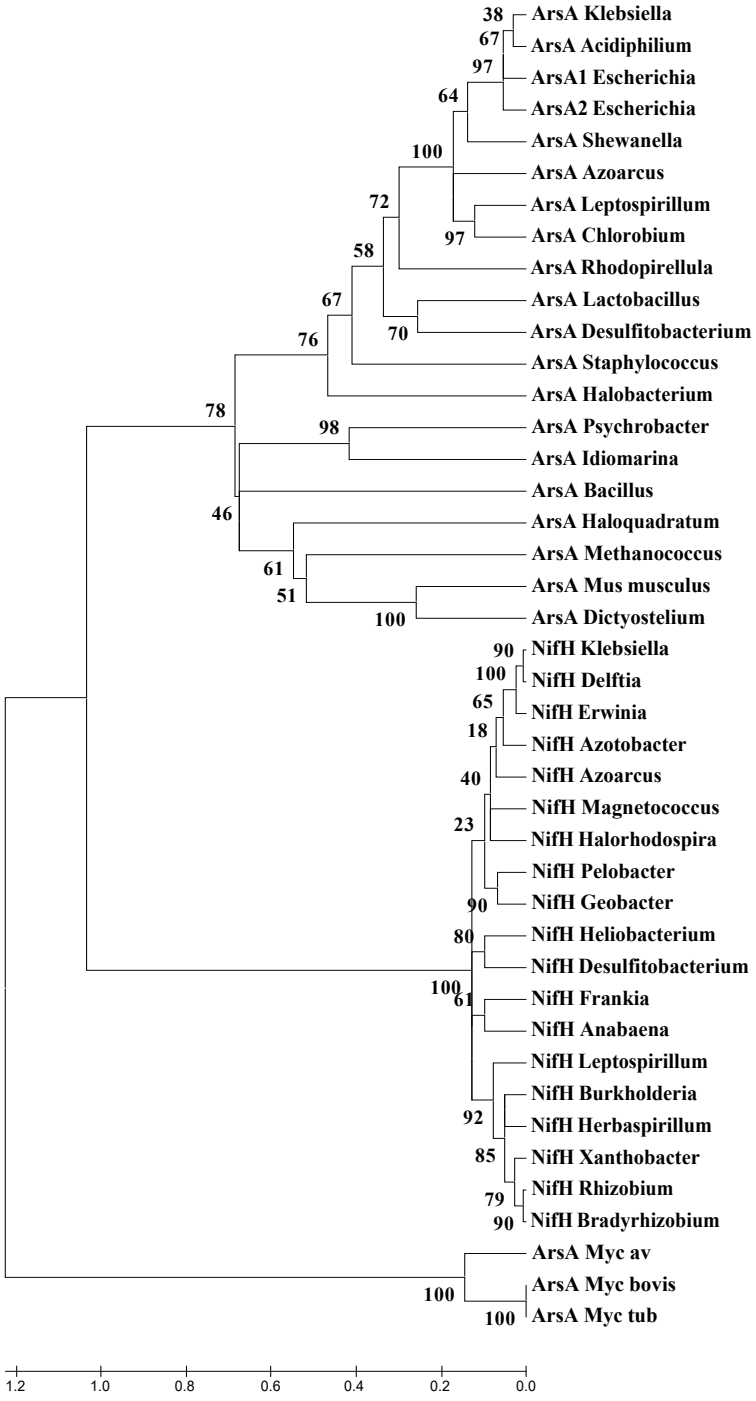
Fig. 5.18 represents a bootstrap consensus phylogenetic tree (500 bootstrap replicates) obtained by the neighbour-joining method from comparison of the NifH and ArsA amino acid sequences. Interestingly, from the phylogenetic analyses, the ArsA protein from the *Mycobacterium* species is shown to be distantly related to the other NifH and ArsA sequences. It may be possible that the *Mycobacterium* ArsA arose much earlier than the other ArsA proteins and the present day ArsA and NifH proteins evolved from such a precursor ArsA protein. It is also evident that whereas large sequence variations have taken place over vast periods of time in the ArsA protein, the NifH proteins have remained closely conserved since the time of their divergence. It is notable

that although the NifH protein is found only in the prokaryotes, the ArsA protein is found in all the domains of life, implying the widespread importance of the extrusion of unconjugated anions from cells.

The various regions of functional importance in ArsA include the cysteines of the allosteric site, aspartates in the Mg2+ binding site, the glycine rich flexible loops and the DTAP sequences in the signal transduction domains (Zhou et al., 2000; Bhattacharjee et al., 1995; Zhou and Rosen, 1999). Although the NifH and ArsA1/ArsA2 amino acid sequences show ~50% similarity to each other, the similarity is more significant in terms of conservation of the secondary structure. The above four important structural features of ArsA are also present in the NifH protein, and may be responsible for the observation of complementation of the function of ArsA1 by NifH. This functional complementation demonstrates the strong conservation of essential domains that have been maintained in two proteins eventhough they diverged to perform varied functions.

Our results clearly define that extant related proteins that have vastly different functions but conserved structural features may verily substitute for the function of one another. Specifically, in the case of the NifH protein, it has been shown before that it could also substitute for the function of the structurally related ChlL protein of *Chlamydomonas reinhardtii* (Cheng et al., 2005). From our studies, it is evident that the NifH can functionally substitute for the function of the ArsA1 protein. Therefore, the NifH protein model is extremely significant in terms of its protein architecture, since it has shown the ability to perform very diverse functions solely on the basis of its conserved structural features. It would be very interesting to investigate the functional

substitution of NifH by ArsA1 or ArsA2 because NifH functions as an obligate electron donor to NifDK for the nitrogen fixation reaction and therefore its functional substitution would be of immense significance in terms of biological nitrogen fixation studies. Finally, the evolutionary aspects of NifH and ArsA indicate that the NifH may have evolved from the ArsA protein and since then strictly maintained its specific sequence, structure and function for nitrogen fixation only. The evolutionary relatedness of ArsA and NifH also leads to the basis of the functional substitution of ArsA1 by NifH because if the NifH has actually emerged from the ancestral ArsA protein, it is more probable that it could still function in arsenite resistance, which is similar to the results of this study.



Percentages from 500 replicate bootstrapping analyses are shown near each branching point.

Fig. 5.18. Neighbour-joining tree based on ArsA and NifH protein phylogenetic analyses.

References

Armstrong, G. A. 1998. Greening in the dark: Light independent chlorophyll biosynthesis from anoxygenic photosynthetic bacteria to gymnosperms. J. Photochem. Photobiol. B 43: 87-100

Bhattacharjee, H., Li, J., Ksenzenko, M. Y., Rosen, B. P. 1995. Role of cysteinyl residues in metalloactivation of the oxyanion-translocating ArsA ATPase. J. Biol. Chem. 270: 11245-11250

Buckel, W. 1980. The reversible dehydratation of (*R*)-2-hydroxyglutarate to (*E*) glutaconate. Eur. J. Biochem. 106: 439-447

Buckel, W. 2003. Archerasen- eine wachsende Enzymfamilie katalysiert ATP-induzierten elektronentransport. BIOspektrum 9: 146-149

Burgess, B. K. and Lowe, D. J. 1996. Mechanism of molybdenum nitrogenase. Chem. Rev. 96: 2983-3011.

Burke, D. H., Hearst, J. E. and Sidow, A. 1993. Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins. Proc. Natl. Acad. Sci. U S A. 90: 7134-7138

Chen, C. M., Misra, T. K., Silver, S. and Rosen BP. 1986. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. J. Biol. Chem. 261: 15030-8

Chen, L., Gavini, N., Tsuruta, H., Eliezer, D., Burgess, B. K., Doniach, S. and Hodgson, K. O. 1994. MgATP-induced conformational changes in the iron protein from *Azotobacter vinelandii*, as studied by small-angle x-ray scattering. J. Biol. Chem. 269(5): 3290-3294.

Cheng, Q., Day, A., Dowson-Day, M., Shen, G-F. and Dixon, R. 2005. The *Klebsiella pneumoniae* nitrogenase Fe protein gene (*nifH*) functionally substitutes for the *chlL* gene in *Chlamydomonas reinhardtii*. Biochem. Biophys. Res. Commun. 329: 966-975

Ching, M. H., Kaur, P., Karkaria, C. E., Steiner, R. F. and Rosen, B.P. 1991. Substrate-induced dimerization of the ArsA protein, the catalytic component of an anion-translocating ATPase. J. Biol. Chem. 266: 2327 - 2332.

Christiansen, J., Chan, J. M., Seefeldt, L. C. and Dean, D. R. 2000. The role of the MoFe protein alpha-125Phe and beta-125Phe residues in *Azotobacter vinelandii* MoFe protein-Fe protein interaction. J. Inorg. Biochem. 80:195-204

Cordell, S. C. and Lowe, J. 2001. Crystal structure of the bacterial cell division regulator MinD. FEBS. Lett. 492: 160-165

de Boer, P. A. J., Crossley, R. E. and Rothfield, L. I. 1992. Roles of MinC and MinD in the site-specific septation block mediated by the MinCDE system of *Escherichia coli*. J. Bacteriol. 174: 63-70

Dey, S., Dou, D., Tisa, L. S. and Rosen, B. P. 1994. Interaction of the catalytic and the membrane subunits of an oxyanion-translocating ATPase. Arch. Biochem. Biophys. 311: 418-24.

Finkelstein, A. V., Gutun, A. M., Badretdinov, A. Y. 1993. Why are the same protein folds used to perform different functions? FEBS Lett. 325: 23-28

Fujita Y, Matsumoto H, Takahashi Y and Matsubara H. 1993. Identification of a *nifDK*-like gene (ORF467) involved in the biosynthesis of chlorophyll in the cyanobacterium *Plectonema boryanum*. Plant Cell Physiol. 34: 305-14

Fujita, Y. 1996. Protochlorophyllide reduction: A key step in the greening of plants. Plants Cell Physiol. 37: 411-421

Fujita, Y., Bauer, C.E. 2000. Reconstitution of light-independent protochlorophyllide reductase from purified Bchl and BchN-BchB subunits. *In vitro* confirmation of nitrogenase like features of a bacteriochlorophyll biosynthesis enzyme. J. Biol.Chem. 275: 23583-23588

Fujita, Y., Takahashi, Y., Kochi, T., Ozeki, H., Ohyama, K. and Matsubara, H. 1989. Identification of a novel NifH-like (frxc) protein in chloroplasts of the Liverwort *Marchantia polymorpha*. Plant. Mol. Biol. 13: 551-561

Gatti, D., Mitra, B. and Rosen, B. P. 2000. *Escherichia coli* soft metal ion-translocating ATPases. J. Biol. Chem. 275: 34009-34012

Gavini, N. and Burgess, B. K. 1992. FeMo cofactor synthesis by a *nifH* mutant with altered MgATP reactivity. J. Biol. Chem. 267: 21179-21186.

Gavini, N. and Pulakat, L. Role of NifM in maturation of the Fe protein of nitrogenase, *In*: T. Finan, M. O'Brian, D. Layzell, K. Vessey, B. Newton (Eds.), Nitrogen Fixation: Global Perspectives, CABI International, Wallingford, Oxford, 2002, pp.228-232

Gavini, N., Ma, L., Watt, G. and Burgess, B. K. 1994. Purification and characterization of a FeMo cofactor-deficient MoFe protein. Biochemistry. 33(39): 11842-11849

Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J. and Rees, D. C. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. Science 1992, 257:1653-1659.

Guex, N. and Peitsch, M. C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 18(15): 2714-2723

Hans, M., Sievers, J., Muller, U., Vorholt, J. A., Linder, D. and Buckel, W. 1999. 2-hydroxyglutaryl-CoA dehydratase from *Clostridium symbiosum*. Eur. J. Biochem. 265: 404-414

Hausinger, R. P. and Howard, J. B. 1983. Thiol reactivity of the nitrogenase Fe-protein from *Azotobacter vinelandii*. J. Biol. Chem. 258: 13486-92.

Hayashi, I., Oyama, T. and Morikawa, K. 2001. Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus. EMBO J. 20(8): 1819-1828

Howard J. B. and Rees, D. C. 1996. Structural basis of biological nitrogen fixation. Chem. Rev. 96: 2965-82

Howard, J. B. and Rees, D. C. 1994. Nitrogenase: a nucleotide-dependent molecular switch. Annu. Rev. Biochem. 63: 235-264

Hu, Z., Gogol, E. P. and Lutkenhaus, J. 2002. Dynamic assembly of MinD on phospholipid vesicles regulated by ATP and MinE. Proc. Natl. Acad. Sci. 99: 6761–6766

Hu, Z., Saez, C. and Lutkenhaus, J. 2003. Recruitment of MinC, an inhibitor of Z - ring formation, to the membrane in *E. coli*: Role of MinD and MinE. J. Bacteriol. 185: 196–203

Huang, C. and Liu, X. Q. 1992. Nucleotide sequence of the *frxC*, *petB* and *trnL* genes in the chloroplast genome of *Chlamydomonas reinhardtii*. Plant Mol. Biol. 18: 985-988

Hurley, J. H., 1996. The sugar kinase/heat shock protein 70/ actin superfamily: implications of conserved structure for mechanism. Annu. Rev. Biophys. Biomol. Struct. 25: 137-162

Jang, S. B., Jeong, M. S., Seefeldt, L. C., Peters, J. W. 2004. Structural and biochemical implications of single amino acid substitutions in the nucleotide-dependent switch regions of the nitrogenase Fe protein from *Azotobacter vinelandii*. J. Biol. Inorg. Chem. 9: 1028-33

Jang, S. B., Seefeldt, L. C., Peters, J .W. 2000. Insights into nucleotide signal transduction in nitrogenase: Structure of an iron protein with MgADP bound. Biochemistry 39: 14745-14752

Kaur, P. and Rosen, B. P. 1993. Complementation between nucleotide binding domains in an anion-translocating ATPase. J. Bacteriol. 175: 351-357.

Kim, J., Hetzel, M., Boiangiu, C. D. and Buckel, W. 2004. Dehydratation of (*R*)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of □-amino acids by anaerobic bacteria. FEMS Microbiol. Rev. 28: 455-468

Koonin, E. V. 1993. A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. J. Mol. Biol. 229: 1165-1174

Kumar, S., Tamura, K. and Nei, M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163

Lanzilotta, W. N., Fisher, K. and Seefeldt, L. C. 1997. Evidence for electron-transfer dependent formation of a nitrogenase iron protein - molybdenum-iron protein tight complex. The role of aspartate 39. J. Biol. Chem. 272: 4157-4165

Lanzilotta, W. N., Ryle, M. J., and Seefeldt, L. C. 1995. Nucleotide hydrolysis and protein conformational changes in *Azotobacter vinelandii* nitrogenase iron protein: defining the function of aspartate 129. Biochemistry 34: 10713-10723

Li, J., Liu, S. and Rosen, B. P. 1996. Interaction of ATP binding sites in the ArsA ATPase, the catalytic subunit of the Ars pump. J. Biol. Chem. 271: 25247-25252

Li, J. and Rosen, B. P. 2000. The linker peptide of the ArsA ATPase. Mol. Microbiol. 35: 361-367

Lill, R. and Muhlenhoff, U. 2005. Iron-sulfur-protein biogenesis in eukaryotes. Trends Biochem. Sci. 30:133-141

Locher, K. P., Hans, M., Yeh, A. P., Schmid, B., Buckel, W. and Rees, D. C. 2001. Crystal structure of the *Acidaminococcus fermentans* 2-hydroxyglutaryl-CoA dehydratase component A. J. Mol. Biol. 307(1): 297-308

Lutkenhaus, J. and Sundaramoorthy, M. 2003. MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. Mol. Microbiol. 48: 295–303

Ma, L., Gavini, N., Liu, H. I., Hedman, B. and Hodgson, K. O., Burgess, B. K. 1994. Large scale isolation and characterization of the molybdenum-iron cluster from nitrogenase. J. Biol. Chem. 269(27): 18007-18015.

Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D. and Bryant, S. H. 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res. 33: D192-196

Milner-White, E. J., Coggins, J. R. and Anton, I. A. 1991. Evidence for an ancestral core structure in nucleotide-binding proteins with the type-A motif. J. Mol. Biol. 221: 751-754

Muller, U. and Buckel, W. 1995. Activation of (*R*)-2-hydroxyglutaryl-CoA dehydratase from *Acidaminococcus fermentas*. Eur. J. Biochem. 230: 698-704

Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247: 536-540

Raskin D. M. and de Boer, P. A. 1999. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. Proc. Natl. Acad. Sci. 96: 4971–4976

Rees, D. C. 2002. Great metalloclusters in enzymology. Ann. Rev. Biochem. 71: 221-246

Rees, D. C., Tezcan, F. A.., Haynes, C. A., Walton, M. Y., Andrade, S., Einsle, O. and Howard, J. B. 2005. Structural basis of biological nitrogen fixation. Philos. Transact. A. Math. Phys. Eng. Sci. 363: 971-984

Rothfield, L. I., Shih, Y.-L. and King, G. 2001. Polar explorers: membrane proteins that determine division site placement. Cell 106: 13–16

Ryle, M. J., Lanzilotta, W. N., Mortenson, L. E., Watt, G. D., and Seefeldt, L. C. 1995. Evidence for a central role of lysine 15 of A*zotobacter vinelandii* nitrogenase iron protein in nucleotide binding and protein conformational changes. J. Biol. Chem. 270: 13112-13117

Sakai, N., Yao, M., Itou, H., Watanabe, N., Yumoto, F. Tanokura, M. and Tanaka, I. 2001. The three-dimensional structure of septum site-determining protein MinD from *Pyrococcus horikoshii* OT3 in complex with MgADP. Structure (*Camb*) 9: 817-826

Sambrook, J. F. Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1992

Schindelin, H., Kisker, C. Schlessman, J. L., Howard, J. B. and Rees, D. C. 1997. Structure of ADP-AlF_4^- stabilized nitrogenase complex and its implications for signal transduction. Nature 387: 370-376

Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B. and Rees, D. C. 1998. Conformational variability in structures of the nitrogenase iron proteins from *Azotobacter vinelandii* and *Clostridium pasteurianum*. J. Mol. Biol. 280: 669-685.

Schmid, B., Einsle, O., Chiu, H. J., Willing, A., Yoshida, M., Howard, J. B. and Rees, D. C. 2002. Biochemical and structural characterization of the cross-linked complex of nitrogenase: comparison to the ADP-AlF_4^--stabilized structure. Biochemistry. 1(52): 15557-15565.

Schulz, G. E. 1992. Binding of nucleotides by proteins. Curr. Opin. Struct. Biol. 2: 61-67

Seefeldt, L. C. and Dean, D. R. 1997. Role of nucleotides in nitrogenase catalysis. Acc. Chem. Res. 30:.260-266

Seefeldt, L. C. and Mortenson, L. E. 1993. Increasing nitrogenase catalytic efficiency for MgATP by changing serine 16 of its Fe protein to threonine: Use of Mn(2+) to show interaction of serine 16 with Mg(2+). Protein Sci. 2: 93-102

Seefeldt, L. C., Morgan, T. V., Dean, D. R. and Mortenson, L. E. 1992. Mapping the site(s) of magnesium-ATP and magnesium-ADP interaction with the nitrogenase of *Azotobacter vinelandii*. Lysine 15 of the iron protein plays a major role in magnesium-ATP interaction. J. Biol. Chem. 267: 6680-6688

Shih, Y-L., Le, T. and Rothfield, L. 2003. Division site selection in *Escherichia coli* involves dynamic redistribution of Min proteins within coiled structures that extend between the two cell poles. Proc. Natl. Acad. Sci. 100**:** 7865–7870.

Sprang, S. R. 1997. G-protein mechanisms: insights from structural analysis. Annu. Rev. Biochem. 66: 639-678

Strop, P., Takahar, P. M., Chiu, H-J., Hayley, C., Angrove, C., Burgess, B. K. and Rees, D. C. 2001. Crystal structure of the all-ferrous $[4Fe\text{-}4S]^0$ form of the nitrogenase iron protein from *Azotobacter vinelandii*. Biochemistry 40: 651-656

Suzuki, J. Y. and Bauer, E. C. 1992. Light-independent chlorophyll biosynthesis: involvement of the chloroplast gene *chlL* (frxC). Plant Cell 4: 929-940

Tal, S., Chun, T. W., Gavini, N. and Burgess, B. K. 1991. The delta *nifB* (or delta *nifE*) FeMo cofactor-deficient MoFe protein is different from the delta *nifH* protein. J. Biol. Chem. 266(16): 10654-10657.

Tezcan, F. A., Kaiser, J. T., Mustafi, D., Walton, M. Y., Howard, J. B and Rees DC. 2005. Nitrogenase complexes: multiple docking sites for a nucleotide switch protein. Science 309:1377-1380.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22(22): 4673-4680

Walker, .J. E., Saraste, M., Runswick, M. J. and Gay, N. J. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 8: 945-981

Wolle, D., Dean, D. R. and Howard, J. B. 1992. Nucleotide-iron-sulfur cluster signal transduction in the nitrogenase iron-protein: the role of Asp125. Science 258: 992-995.

Wurch, T., Lestienne, F., Pauwels, P. J. 1998. A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. Biotech. Techniq. 12: 653-657.

Zhou, T., Radaev, S., Rosen, B. P. and Gatti, D. L. 2001. Conformational changes in four regions of the *Escherichia coli* ArsA ATPase link ATP hydrolysis to ion translocation. J. Biol. Chem. 276: 30414-30422

Zhou, T., Radaev, S., Rosen, B. P. and Gatti, D. L. 2000. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. EMBO J. 19: 4838-4845

Zhou, T. and Rosen, B. P. 1999. Asp45 is a Mg2+ ligand in the ArsA ATPase. J. Biol. Chem. 274: 13854-13858

CHAPTER VI

SUMMARY

The nitrogenase enzyme that catalyzes the conversion of nitrogen (N) to ammonia is one of the most intriguing and important enzymes, vital for the maintenance of life on earth. Only a selected group of microorganisms known as diazotrophs possess the nitrogenase enzyme and therefore have the ability to fix nitrogen. The source of N for crops is provided either by chemical N fertilizers or by the fixed nitrogen from symbiotic or free-living diazotrophic activity. The prospect of extending the ability of nitrogen fixation to plants in order to remove N nutrition limitations has remained a challenging goal for scientists. The options include encouraging the development of symbiosis between other plant and bacterial pairs or transfer of the nitrogen fixation genes to the plants by genetic engineering (Aldridge, 1996). However about 17 different genes are involved in the nitrogen fixation process and since multi-gene transfer is a difficult task, one of the first issues to be addressed is the identification of the minimal set of required genes and their compression. Also, it is important to find out the function of every *nif* gene encoded product and their biological networks.

Research on nitrogenase has provided clues to its structure, organization, mechanism and function. Nitrogenase is a complex metalloenzyme and major emphasis has therefore been laid on its genetic organization, maturation and assembly of its metalloclusters via the FeMoco biosynthetic pathway, its oxygen sensitivity and also its

structural homology with other metalloenzymes involved in diverse functions. With the advent of molecular biology and bioinformatics, new tools such as mutagenesis, global gene expression, structural homology modeling and phylogenetic analyses are being widely used to solve questions related to this enzyme (examples: Abed et al., 2006; Parro and Moreno-Paz, 2003, Perry et al., 2005, Moisander et al., 2006, Starker et al., 2006). Using many of these tools, we have directed our study towards the NifHDKX proteins and obtained new insights regarding their protein interactions and their structural and evolutionary features.

A fused NifDK protein was previously constructed in our lab as a step towards achieving the goal of minimizing the size of the *nif* genes so as to facilitate their transfer into plants. As discussed in Chapter II of this dissertation, we determined protein-protein interactions between the NifD, NifK and NifDK fusion protein to understand the functioning of the fusion protein. It was important to determine whether this fused NifDK protein functioned as a homodimer equivalent to the native heterotetrameric MoFe protein or differently. We utilized the BacterioMatch™ Two Hybrid System to detect these protein-protein interactions. It was found that the homodimers of the NifDK fusion proteins functioned similarly to the heterotetrameric MoFe protein, in terms of the interaction at the â-â (NifK/NifK) interface of the nitrogenase enzyme. Future studies may utilize the *A. vinelandii* strain containing the fusion NifDK protein and study the effect of disruption of the â-â interface for obtaining a single functional □□ (NifDK) unit.

The ChlBN proteins of the protochlorophyllide reductase system bear close sequence similarity to the NifDK proteins of the nitrogenase enzyme system. We

therefore used structural homology modeling tools to investigate if the FeMoco metallic cluster found in the NifD protein could also be present in the ChlN protein. As described in Chapter III, the NifD structure (from the Protein Data Bank) was used as a structural template for homology modeling of the ChlN structure. We examined the derived ChlN structure for the presence of putative FeMoco ligands and also searched for corresponding residues that are important in the FeMoco surroundings in NifD. The possibility of the presence of a FeMoco was strengthened by our finding that the Cys275 and His442 ligands of the FeMoco in NifD corresponded to the Cys326 and His515 residues of the ChlN protein. We also highlighted other important residues in ChlN that could be the counterparts of the residues that surround the FeMoco in NifD. Finally, a phylogenetic analysis of the NifDKEN and ChlBN proteins was performed to understand their evolutionary interrelationships. Future mutagenesis studies directed towards alteration of the Cys326 and His515 residues of ChlN could lead to the verification of our structural studies.

The FeMoco present in the MoFe (NifDK) protein is known as the substrate reduction site of the nitrogenase enzyme. The biosynthesis and insertion of the FeMoco involves several *nif*/*naf* gene products such as NifE, NifN, NifB, NifH, NifX, NifV, NifQ and NafY. We were interested in determining the role of the comparatively less studied NifX in FeMoco biosynthesis. We therefore investigated the protein-protein interactions of NifX with NifB, NifN, NifH, NifD, NifK and NafY using a bacterial two hybrid system. As explained in Chapter IV, the protein-protein interactions detected

between NifX and other Nif proteins allowed us to propose a modified model for the role of NifX in the FeMoco biosynthetic pathway.

In Chapter V, the NifH and its structural homologs (ChlL, CompA, MinD and ArsA) have been reviewed. The functional effect of a novel NifH-ArsA2 chimeric protein on arsenite resistance has been described. The NifH-ArsA2 protein was constructed by using a double PCR method (Wurch et al., 1998) and tested for complementation of an arsenite sensitive *E. coli* strain. The results established that the NifH-ArsA2 protein could functionally substitute for the ArsA protein. It was evident that the conservation of the major structural features of the ArsA1 protein in NifH was sufficient to make it function as the ArsA1 protein. Also phylogenetic data indicated that the ArsA protein has diverged largely during evolution whereas the NifH has remained highly conserved. Very importantly, this revealed that although NifH is highly oxygen sensitive, it could function in presence of oxygen when expressed in a different codon context, such as with the ArsA2 protein.

In conclusion, our study has offered new perspectives in structural, evolutionary and functional aspects of the nitrogenase enzyme. The most important outcomes of this study are the interactions detected between two fused dimeric NifDK proteins, the finding of putative FeMoco ligands in the structurally homologous ChlN protein, a novel role of the NifX protein in the FeMoco biosynthetic pathway and the functional complementation of the ArsA protein by a chimeric NifH-ArsA2 protein. Using microarray studies, we have also provided preliminary inferences for the difference in gene expression related to energy metabolism between the nitrogen-fixing *Azotobacter*

vinelandii and the closely related *Pseudomonas aeruginosa* bacterium so as to understand the respiratory protection of the oxygen sensitive nitrogenase system in *Azotobacter vinelandii.*

This study is of great impact in understanding the structure-function relationship of nitrogenases and other structurally similar enzymes and furthering research related to its interactive molecular network.

References

Abed, R. M., Palinska, K. A., Camoin, G. and Golubic, S. 2006. Common evolutionary origin of planktonic and benthic nitrogen-fixing oscillatoriacean cyanobacteria from tropical oceans. FEMS Microbiol. Lett. 260(2): 171-177

Aldridge, S. The thread of life: The story of genes and genetic engineering. 1996. pp. 206, Cambridge University Press, Cambridge (England); New York, NY, USA

Moisander, P. H., Shiue, L., Steward, G. F., Jenkins, B. D., Bebout, B. M. and Zehr, J. P. 2006. Application of a *nifH* oligonucleotide microarray for profiling diversity of N_2-fixing microorganisms in marine microbial mats. 8(10): 1721-1735

Parro, V., Moreno-Paz, M. 2003. Gene function analysis in environmental isolates: the *nif* regulon of the strict iron-oxidizing bacterium *Leptospirillum ferrooxidans*. Proc. Natl. Acad. Sci. USA 100(13): 7883-7888

Perry, S., Shearer, N., Little, R. and Dixon, R. 2005. Mutational analysis of the nucleotide-binding domain of the anti-activator NifL. J. Mol. Biol. 346(4): 935-949

Starker, C. G., Parra-Colmenares, A. L., Smith, L., Mitra, R. M. and Long, S. R. 2006. Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. Plant. Physiol. 140(2): 671-680

Wurch, T., Lestienne, F., Pauwels, P. J. 1998. A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. Biotech. Techniq. 12: 653-657

APPENDIX A

COMPARATIVE TRANSCRIPTIONAL PROFILE OF THE GENES INVOLVED IN ENERGY METABOLISM IN *AZOTOBACTER VINELANDII* VS *PSEUDOMONAS AERUGINOSA* PAO1

Introduction

Azotobacter vinelandii is a widely distributed free-living soil bacterium. It is Gram-negative, strictly aerobic and possesses several interesting aspects for research, such as the ability to grow on a wide variety of carbohydrates, alcohols and organic acids, alginate production and nitrogen fixation (www.azotobacter.org). The nitrogenase enzyme that catalyzes biological nitrogen fixation is composed of two subunits, the Iron protein (Fe) and the Molybdenum-Iron protein (Mo-Fe) (Kim and Rees, 1992; Georgiadis et al., 1992). This process requires reducing equivalents as well as the supply of atleast 16 ATP per N_2 fixed. Different types of nitrogenase complexes have been described (Molybdenum-Iron, Vanadium, Iron) based on their metal centers and are known as the Nif, Vnf and Anf systems respectively. *In vitro*, nitrogenases are O_2 sensitive and thus most diazotrophs are unable to fix nitrogen under the presence of atmospheric oxygen concentrations. However, *A. vinelandii*, owing to its high respiration rate, can protect its nitrogenase enzyme from oxygen inactivation and thus fix nitrogen under aerobic conditions (http://www.azotobacter.org). It is believed that the high rate of O_2 consumption characteristic of *Azotobacter* species growing at higher O_2 concentrations is excess of the actual energy requirements. However since *Azotobacters* change the composition of the respiratory chain and the coupling of electron transport to ATP regeneration, this problem is overcome. Thus, the respiratory chain composition and regulation has been largely focused upon in previous studies (Postgate, 1998; Poole and Hill, 1997; Hill, 1992). The other processes suggested to be involved in protection of this

enzyme from oxygen include the reversible conversion of the enzyme into a protected inactivated state and the formation of an effective O_2 barrier due to the production of an alginate capsule on the cell surface (Linkerhägner and Oelze, 1995; Liu et al., 1995; Moshiri et al., 1995; Sabra et al., 2000). Several other speculations have been made regarding this intricate challenge of maintaining an intracellular atmosphere wherein the nitrogenase enzyme is protected from the deleterious effect of O_2 and yet a constant supply of O_2 required for energy regeneration during aerobic respiration is available (Gallon, 1992; Bergman et al., 1997).

Although previous studies have provided important clues about the protection of the nitrogenase enzyme in *A. vinelandii* due to its high respiration rate (Linkerhägner and Oelze, 1995; Liu et al., 1995; Moshiri et al., 1995), there has been no further investigation into its corroboration at the molecular level. We were therefore interested in examining the transcriptional profile of the *A. vinelandii* energy metabolism / respiration related genes (energy-metabolism) and determining if there was a higher expression of certain genes compared to that of a closely related bacterium, *Pseudomonas aeruginosa* PAO1 (Rediers et al., 2000). Generation of phylogenetic trees for many housekeeping genes has revealed that the *A. vinelandii* homologues clustered within, or close to, the P*seudomonas* clade (Rediers et al., 2004). In most cases the *A.vinelandii* proteins were most closely related to the *P. aeruginosa* PAO1 orthologues (Rediers et al., 2004). The *Azotobacter* Genome Sequencing Project is in its final stages of annotation and currently only the entire draft sequence of its DNA is available in the GenBank database. Cross-species hybridization has been previously utilized for expression

profiling of organisms that do not have an extant microarray platform (Renn et al., 2004). On account of the high DNA sequence similarity between *A. vinelandii* and *P. aeruginosa*, we utilized the readily available Affymetrix® *P. aeruginosa* DNA chips to hybridize the cDNA prepared from *A. vinelandii* cells grown in Burke's Nitrogen medium (Strandberg et al., 2004).

Materials and Methods

Bacterial strains and growth. Pseudomonas aeruginosa PAO1 was obtained upon request from the University of Washington Genome Center, Seattle (WA). *Azotobacter vinelandii* (AvOP wild type strain) was taken from the laboratory stock. Both strains were cultured in BN+ media that contains ammonium acetate as a nitrogen source (Strandberg et al., 1968).

Transcriptional analysis. The Affymetrix® PAO1 microarrays that represent the annotated genome of *P. aeruginosa* strain PAO1 (containing 25-mer probe sets for 5,549 protein-coding sequences) were obtained from Affymetrix® (Santa Clara, CA). Each chip includes18 tRNA genes, a representative of the ribosomal RNA cluster, 117 genes present in *P. aeruginosa* strains other than PAO1 and 199 probe sets corresponding to all intergenic regions exceeding 600 base pairs. Each target transcript is queried by a probe set of the gene chip and each probe set is comprised of a collection of probe pairs that interrogate the same sequence. A probe set typically consists of 11-20 probe pairs. Each probe of a probe pair is designed to detect a perfect match (PM) and a mismatch (MM). The PM probe is a 25-mer oligonucleotide designed to be complementary to a reference sequence and the MM probe is a 25-mer oligonucleotide designed to be complementary

to a reference sequence except for a single nucleotide at the 13th position. The MM probe serves as a measurement of non-specific hybridization. A qualitative measurement indicating if a given transcript is detected (Present (P)), not detected (Absent (A)) or marginally detected (Marginal (M)) is calculated using the discrimination score which is a relative difference between a PM and its MM (R=PM-MM/PM+MM). A complete description and annotation for the *P. aeruginosa* genome array is available at http://www.affymetrix.com.

Sample preparation. The PAO1 and AvOP strains were inoculated in BN+ media (Strandberg et al., 1968) and grown for 18 h. Cultures were diluted 1: 50 and grown until the $O.D._{600}$ ~ was 0.6-1.0). 1ml from this culture was harvested for RNA isolation. Total RNA was extracted using the SV Total RNA Isolation System (Promega, WI), according to the manufacturer's protocol (Promega Technical Manual No. 048, Promega, WI). Before cDNA synthesis, Poly-A RNA controls (*B. subtilis* genes *lys, phe, thr, dap)* were spiked into the RNA and carried through the sample preparation process and evaluated as internal control genes. The final concentrations of the *lys, phe, thr and dap* controls relative to the total RNA populations were 1:100,000, 1:50,000, 1:25,000 and 1:7,500 respectively. We then used 10 □g of total RNA with random primers and SuperScript II (Invitrogen Corp., Carlsbad, CA) for cDNA synthesis, cDNA fragmentation, labeling and hybridization. Sample hybridization efficiency was evaluated using a spiked in hybridization cocktail that contained a mixture of biotin-labeled cRNA transcripts of *bioB*, *bioC*, *bioD* and *cre* in final concentrations of 1.5 pM, 5 pM, 25 pM and 100 pM respectively. These and subsequent staining and washing steps were performed according

to the manufacturer's protocol for the Affymetrix® *P. aeruginosa* GeneChip arrays (Affymetrix®, Inc., Santa Clara, CA). The arrays were scanned with the Affymetrix® Gene Chip Scanner 3000. The hybridization and scanning steps were performed by Ms. Juliet Tang (Facilities Manager) at the Life Sciences and Biotechnology Institute (LSBI), Mississippi State University, MS. Finally, using four *P. aeruginosa* genome chips, gene expression data was obtained from two biological replicates of *P. aeruginosa* and two biological replicates of *A. vinelandii*.

Data analysis. The data obtained was analyzed by using the Affymetrix Gene Chip Operating Software (GCOS) v 1.0 and GeneTraffic (Stratagene, La Jolla, CA) software. The fold change was calculated as the ratio between the signal averages of two *P. aeruginosa* and two *A. vinelandii* cultures. For same species microarray studies, data is usually sorted based on the presence of three criteria: a 'Present' call detection indicating reliable detection of the transcript, p-value of 0.05 and a fold change of 2. As stated earlier, the 'Present' call is assigned after considering both, PM and the MM values. However, in case of cross-species hybridization such as this between an *A. vinelandii* target and a *P. aeruginosa* probe, including the MM factor could reduce the number of genes obtained as being highly expressed because even slight mismatch of sequence would lead to an 'Absent' call. Therefore, we have sorted the data based on the presence of the following criteria: (1) p-value significance of 0.05 (indicating minimal error and statistical significance) and (2) a fold change of atleast 2.

Results and Discussion

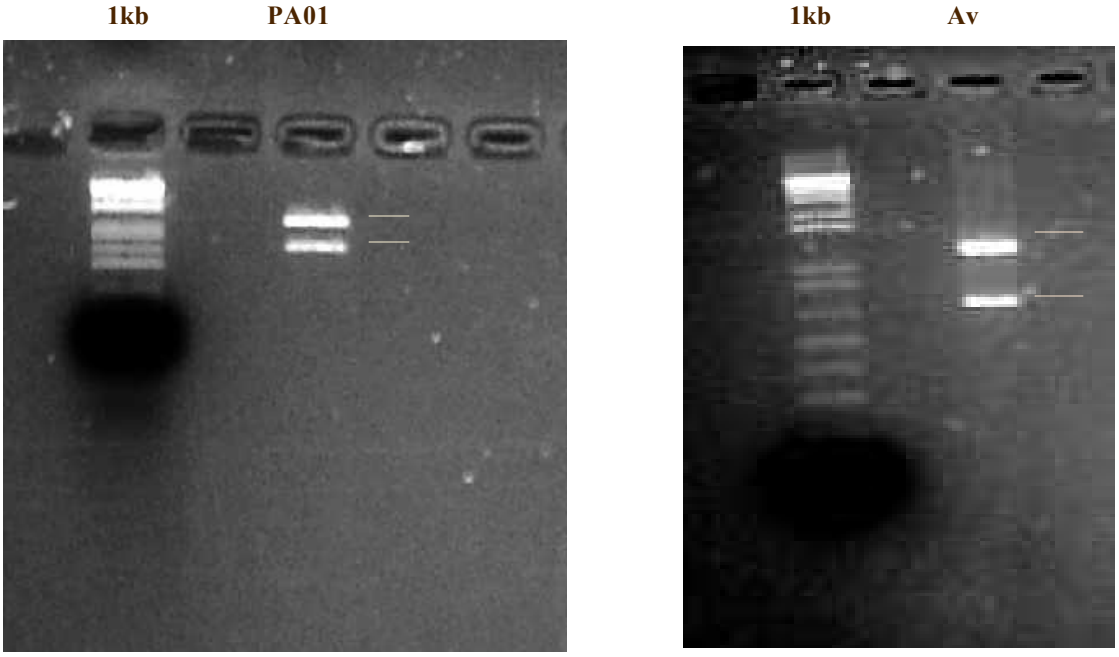
As shown in Table A.1, the general distribution of the gene products in *P. aeruginosa* and *A. vinelandii* was compared according to functional categories, using genomic information obtained from www.microbesonline.org.

Table A.1. Comparison of the distribution of gene products according to functional categories in *Azotobacter vinelandii* (*Av*) and *Pseudomonas aeruginosa* (*Pae*) (www.microbesonline.org)

Function	***Av***	***Pae***
A: RNA processing and modification	-	2
B: Chromatin structure and dynamics	1	3
C: Energy production and conversion	323	311
D: Cell division and chromosome partitioning	37	34
E: Amino acid transport and metabolism	281	491
F: Nucleotide transport and metabolism	78	106
G: Carbohydrate transport and metabolism	248	207
H: Coenzyme metabolism	158	191
I: Lipid metabolism	134	191
J: Translation, ribosomal structure and biogenesis	169	197
K: Transcription	277	474
L: DNA replication, recombination, and repair	194	133
M: Cell envelope biogenesis, outer membrane	189	245
N: Cell motility and secretion	77	149
O: Posttranslational modification, protein turnover, chaperones	159	184
P: Inorganic ion transport and metabolism	279	304
Q: Secondary metabolites biosynthesis, transport, and catabolism	125	157
R: General function prediction only	417	571
S: Function unknown	338	482
T: Signal transduction mechanisms	189	327
U: Intracellular trafficking and secretion	106	165
V: Defense mechanisms	47	75

Accordingly, a total of 323 proteins are related to energy metabolism processes in *A. vinelandii* compared to 311 in *P. aeruginosa*. Also, the functions of carbohydrate transport and metabolism and DNA replication, recombination and repair are found to recruit many more proteins in *A. vinelandii* than *P. aeruginosa*.

The quality of the total RNA extracted from the *P. aeruginosa* and *A. vinelandii* cultures was examined on an agarose gel prior to starting the assay. As shown in Fig. A.1 and A.2, the 23S and 16S rRNA bands of both extracted RNA samples were seen clearly and no smearing was observed, indicating good purity of the samples.

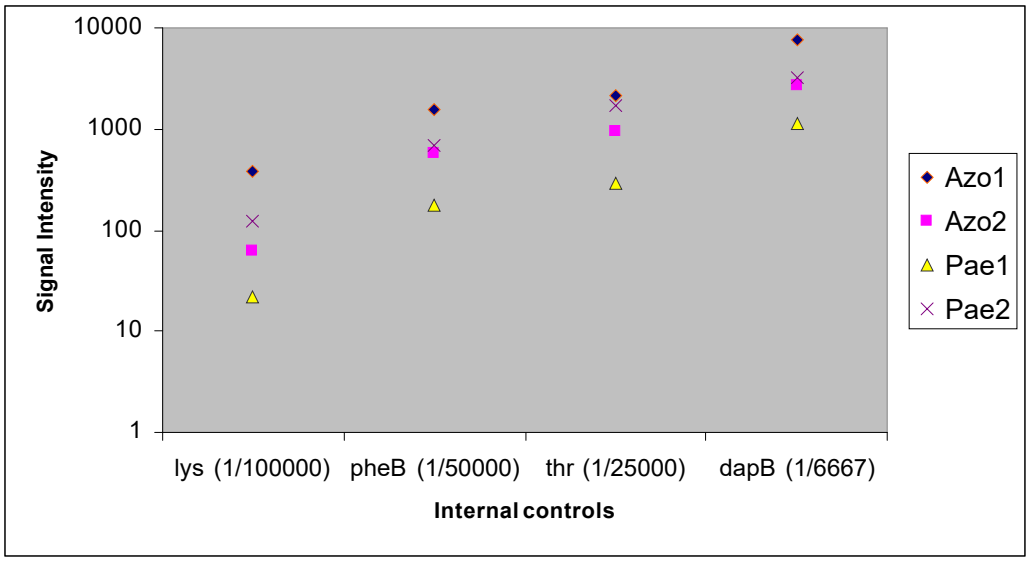


The arrows indicate the upper 23S rRNA band and the lower 16S rRNA band and also indicate good general purity of the RNA sample.

Fig. A.1 (left) Purity of RNA from *P. aeruginosa* culture and

Fig. A.2 (right) Purity of RNA from *A. vinelandii* culture

After obtaining the gene chip data, we investigated the signal intensities of the internal Poly-A RNA controls which were spiked into the total RNA and used for further processing for final hybridization. As shown in Fig. A.3, signal intensities for the *lys, phe, thr and dap* controls (final concentrations relative to total RNA population: 1:100,000, 1:50,000, 1:25,000 and 1:7500 respectively) increased in the same order (see Table A.2. supplementary material). This indicated that the cDNA synthesis, labeling and fragmentation steps were performed accurately.



The *lys, phe, thr and dap* controls were spiked into the total RNA sample. Final concentrations relative to total RNA population were: 1:100,000, 1:50,000, 1:25,000 and 1:6667 respectively. Increase in signal intensities of the transcripts was detected in the same order as their initial concentrations. Azo1 and Azo2 are the gene chip experiments performed using the *A. vinelandii* cDNA and Pae1 and Pae2 are those performed with *P. aeruginosa* cDNA.

Fig. A.3. Signal intensity values for Poly-A RNA controls.

We then analyzed the data obtained by calculating the percentage genes detected as having a 'Present' call in *A. vinelandii* and *P. aeruginosa* and sorted them according to different functional classes based on their gene function description. Fig. A.4 shows an overview of the comparison. On an average, the total percentage of genes shown to have a 'Present' call was found to be ~ 43% for *P. aeruginosa* and ~15% for *A. vinelandii*. This large inequality in expressed gene numbers between *P. aeruginosa* and *A. vinelandii* could be attributed to the expected non-specificity due to cross-hybridization. Table A. 3 (supplementary material) shows the complete set of genes in the *P. aeruginosa* gene chip that are described as having energy metabolism related functions and their respective signal intensities, fold change, p-value and call detection in comparison to *A. vinelandii*. The various sub-groups within the energy metabolism group of genes include ATP synthases, quinones, cytochromes, cofactor biosynthesis, amino acid biosynthesis and carbon compound catabolism related genes. Table A. 4 (supplementary material) shows the fold change, call detection and p-values of these sub-groups. The most significant genes from the entire list of energy metabolism genes comprising of those that show a fold change of 2 and p-value of 0.05 are listed in Table A. 5 (supplementary material).

When the energy metabolism genes are sorted by the above criteria, a list of 12 genes is obtained (Table A. 6). A detailed description of the functions of these genes is given in Table A. 6.

Table A.2. Internal controls (*P. aeruginosa* microarray)

	Azo 1	Azo 2	Pae 1	Pae 2			Azo	Pae					
Gene	**Signal**	**Signal**	**Signal**	**Signal**	**Fold change**	**p-value**	**Detection**	**Detection**	**Gene description**				
AFFX-Bsubtilis_dapB_at	2728.8	7558.5	3215.4	1143.5	2.4	0.3764	P	P	L38424 Bacillus subtilis dapB, dihydropicolinate reductase, nucleotides 1358-2101				
AFFX-Bsubtilis_lys_at	62.9	385.8	121.5	22	3.0	0.4617	P	P	X17013 Bacillus subtilis lys, diaminopimelate decarboxylase, nucleotides 362-1345				
AFFX-Bsubtilis_pheB_at	586.2	1586.3	680.6	180.1	2.8	0.3616	P	P	M24537 Bacillus subtilis pheB, phenylalanine biosynthesis associated protein				
AFFX-Bsubtilis_thrC_at	946.1	2163.8	1690.5	295.8	2.0	0.6057	P	P	X04603 Bacillus subtilis thrC, threonine synthase, nucleotides 248-1306				
Azo1, Azo 2= Microarray experiments utilizing cDNA from *Azotobacter vinelandii*													
Pae1, Pae 2= Microarray experiments utilizing cDNA from *Pseduomonas aeruginosa*													

Table A.3. List of energy metabolism related genes

Fold change	p-value	Av	Pae	Gene description
1.8	0.5057	A	P	PA0023 /GENE=qor /DEF=quinone oxidoreductase /FUNCTION=Energy metabolism
0.4	0.4580	P	P	PA0105 /GENE=coxB /DEF=cytochrome c oxidase, subunit II /FUNCTION=Energy metabolism
3.6	0.5614	A	P	PA0106 /GENE=coxA /DEF=cytochrome c oxidase, subunit I /FUNCTION=Energy metabolism
0.3	0.4296	A	P	PA0107 /DEF=conserved hypothetical protein /FUNCTION=Energy metabolism
0.4	0.4450	P	P	PA0108 /GENE=coIII /DEF=cytochrome c oxidase, subunit III /FUNCTION=Energy metabolism
1.6	0.6586	P	P	PA0113 /DEF=probable cytochrome c oxidase assembly factor /FUNCTION=Energy metabolism
0.6	0.5377	A	P	PA0195 /GENE=pntA /DEF=still frameshift pyridine nucleotide transhydrogenase alpha subunit /FUNCTION=Energy metabolism; Transport of small mo
2.1	0.4404	A	P	PA0196 /GENE=pntB /DEF=pyridine nucleotide transhydrogenase, beta subunit /FUNCTION=Energy metabolism; Transport of small molecules
5.0	0.0306	P	P	PA0330 /GENE=rpiA /DEF=ribose 5-phosphate isomerase /FUNCTION=Energy metabolism
0.9	0.7322	P	P	PA0362 /GENE=fdx1 /DEF=ferredoxin [4Fe-4S] /FUNCTION=Energy metabolism
2.3	0.9963	P	P	PA0509 /GENE=nirN /DEF=probable c-type cytochrome /FUNCTION=Energy metabolism; Biosynthesis of cofactors, prosthetic groups and carriers
0.2	0.4219	A	P	PA0510 /DEF=probable uroporphyrin-III c-methyltransferase /FUNCTION=Energy metabolism; Biosynthesis of cofactors, prosthetic groups and carriers
2.5	0.9376	A	P	PA0511 /GENE=nirJ /DEF=heme d1 biosynthesis protein NirJ /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabolism
0.2	0.4257	A	P	PA0512 /DEF=conserved hypothetical protein /FUNCTION=Energy metabolism; Hypothetical, unclassified, unknown; Biosynthesis of cofactors, prosthet
2.1	0.9481	A	P	PA0513 /DEF=probable transcriptional regulator /FUNCTION=Energy metabolism; Transcriptional regulators; Biosynthesis of cofactors, prosthetic g
0.3	0.4466	A	P	PA0514 /GENE=nirL /DEF=heme d1 biosynthesis protein NirL /FUNCTION=Energy metabolism; Hypothetical, unclassified, unknown;
0.1	0.4076	A	P	PA0515 /DEF=probable transcriptional regulator /FUNCTION=Energy metabolism; Transcriptional regulators; Biosynthesis of cofactors, prosthetic gr
0.7	0.4656	A	P	PA0516 /GENE=nirF /DEF=heme d1 biosynthesis protein NirF /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabol
0.1	0.4249	A	P	PA0517 /GENE=nirC /DEF=probable c-type cytochrome precursor /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy m
0.4	0.4390	A	P	PA0518 /GENE=nirM /DEF=cytochrome c-551 precursor /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabolism
0.4	0.4355	A	P	PA0519 /GENE=nirS /DEF=nitrite reductase precursor /FUNCTION=Energy metabolism
1.3	0.8499	A	P	PA0520 /GENE=nirQ /DEF=regulatory protein NirQ /FUNCTION=Central intermediary metabolism; Energy metabolism
3.1	0.3772	A	P	PA0521 /DEF=probable cytochrome c oxidase subunit /FUNCTION=Energy metabolism
0.3	0.4474	A	P	PA0523 /GENE=norC /DEF=nitric-oxide reductase subunit C /FUNCTION=Energy metabolism
0.4	0.4635	A	P	PA0524 /GENE=norB /DEF=nitric-oxide reductase subunit B /FUNCTION=Energy metabolism
1.5	0.8567	P	P	PA0525 /DEF=probable dinitrification protein NorD /FUNCTION=Energy metabolism

Table A.3. (contd.)

9.9	0.2432	A	P	PA0552 /GENE=pgk /DEF=phosphoglycerate kinase /FUNCTION=Carbon compound catabolism; Energy metabolism
0.2	0.1247	A	P	PA0589 /DEF=conserved hypothetical protein /FUNCTION=Energy metabolism
13.1	0.0647	A	P	PA0607 /GENE=rpe /DEF=ribulose-phosphate 3-epimerase /FUNCTION=Energy metabolism
1.8	0.0737	A	P	PA0609 /GENE=trpE /DEF=anthranilate synthetase component I /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism
2.0	0.6518	A	P	PA0649 /GENE=trpG /DEF=anthranilate synthase component II /FUNCTION=Amino acid biosynthesis and metabolism;
3.0	0.3509	A	P	PA0794 /DEF=probable aconitate hydratase /FUNCTION=Energy metabolism
0.8	0.9822	A	P	PA0854 /GENE=fumC2 /DEF=fumarate hydratase /FUNCTION=Carbon compound catabolism; Energy metabolism
0.3	0.0470	A	P	PA0918 /DEF=cytochrome b561 /FUNCTION=Energy metabolism
3.4	0.4269	A	P	PA0927 /GENE=ldhA /DEF=D-lactate dehydrogenase (fermentative) /FUNCTION=Carbon compound catabolism; Central intermediary metabolism;
2.0	0.2721	A	P	PA1104 /GENE=fliI /DEF=flagellum-specific ATP synthase FliI /FUNCTION=Motility & Attachment; Energy metabolism
0.3	0.3847	A	P	PA1172 /GENE=napC /DEF=cytochrome c-type protein NapC /FUNCTION=Energy metabolism
0.6	0.5562	P	P	PA1173 /GENE=napB /DEF=cytochrome c-type protein NapB precursor /FUNCTION=Energy metabolism
0.3	0.4463	A	P	PA1174 /GENE=napA /DEF=periplasmic nitrate reductase protein NapA /FUNCTION=Energy metabolism
0.7	0.6005	P	P	PA1175 /GENE=napD /DEF=NapD protein of periplasmic nitrate reductase /FUNCTION=Energy metabolism
0.1	0.3218	A	P	PA1176 /GENE=napF /DEF=ferredoxin protein NapF /FUNCTION=Energy metabolism
0.1	0.2668	A	P	PA1177 /GENE=napE /DEF=periplasmic nitrate reductase protein NapE /FUNCTION=Energy metabolism
7.5	0.3583	A	A	PA1317 /GENE=cyoA /DEF=cytochrome o ubiquinol oxidase subunit II /FUNCTION=Energy metabolism
6.2	0.0444	A	A	PA1318 /GENE=cyoB /DEF=cytochrome o ubiquinol oxidase subunit I /FUNCTION=Energy metabolism
6.0	0.2050	A	A	PA1319 /GENE=cyoC /DEF=cytochrome o ubiquinol oxidase subunit III /FUNCTION=Energy metabolism
3.1	0.5212	A	A	PA1320 /GENE=cyoD /DEF=cytochrome o ubiquinol oxidase subunit IV /FUNCTION=Energy metabolism
4.8	0.3794	A	A	PA1321 /GENE=cyoE /DEF=cytochrome o ubiquinol oxidase protein CyoE /FUNCTION=Energy metabolism
22.1	0.0698	A	A	PA1400 /DEF=probable pyruvate carboxylase /FUNCTION=Energy metabolism
2.5	0.5341	P	P	PA1479 /GENE=ccmE /DEF=cytochrome C-type biogenesis protein CcmE /FUNCTION=Energy metabolism
1.8	0.9527	A	P	PA1480 /GENE=ccmF /DEF=cytochrome C-type biogenesis protein CcmF /FUNCTION=Energy metabolism
1.7	0.7886	A	P	PA1481 /GENE=ccmG /DEF=cytochrome C biogenesis protein CcmG /FUNCTION=Chaperones & heat shock proteins; Energy metabolism

Table A.3. (contd.)

1.3	0.7097	P	P	PA1482 /GENE=ccmH /DEF=cytochrome C-type biogenesis protein CcmH /FUNCTION=Energy metabolism
0.3	0.4512	A	P	PA1483 /GENE=cycH /DEF=cytochrome c-type biogenesis protein /FUNCTION=Energy metabolism
2.0	0.5171	A	A	PA1498 /GENE=pykF /DEF=pyruvate kinase I /FUNCTION=Carbon compound catabolism; Energy metabolism
0.2	0.4367	A	P	PA1551 /DEF=probable ferredoxin /FUNCTION=Energy metabolism
0.6	0.6999	A	P	PA1552 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
0.2	0.3484	A	P	PA1553 /DEF=probable cytochrome c oxidase subunit /FUNCTION=Energy metabolism
0.1	0.3699	A	P	PA1583 /GENE=sdhA /DEF=succinate dehydrogenase (A subunit) /FUNCTION=Energy metabolism
0.5	0.4750	A	P	PA1584 /GENE=sdhB /DEF=succinate dehydrogenase (B subunit) /FUNCTION=Energy metabolism
0.2	0.3902	A	P	PA1585 /GENE=sucA /DEF=2-oxoglutarate dehydrogenase (E1 subunit) /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism
2.5	0.8099	P	P	PA1586 /GENE=sucB /DEF=dihydrolipoamide succinyltransferase (E2 subunit) /FUNCTION=Energy metabolism
0.1	0.3547	A	P	PA1587 /GENE=lpdG /DEF=lipoamide dehydrogenase-glc /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism
0.2	0.4196	A	P	PA1588 /GENE=sucC /DEF=succinyl-CoA synthetase beta chain /FUNCTION=Energy metabolism
1.1	0.6946	P	P	PA1589 /GENE=sucD /DEF=succinyl-CoA synthetase alpha chain /FUNCTION=Energy metabolism
2.5	0.2962	P	A	PA1600 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
0.2	0.4043	A	P	PA1770 /GENE=ppsA /DEF=phosphoenolpyruvate synthase /FUNCTION=Carbon compound catabolism; Central intermediary metabolism; Energy metab
1.3	0.8506	A	P	PA1787 /GENE=acnB /DEF=aconitate hydratase 2 /FUNCTION=Energy metabolism
4.1	0.2119	A	A	PA1883 /DEF=probable NADH-ubiquinone/plastoquinone oxidoreductase /FUNCTION=Energy metabolism
1.6	0.8346	A	P	PA1931 /DEF=probable ferredoxin /FUNCTION=Carbon compound catabolism; Energy metabolism
1.1	0.9484	A	P	PA1983 /GENE=exaB /DEF=cytochrome c550 /FUNCTION=Carbon compound catabolism; Energy metabolism
8.1	0.0021	A	A	PA2153 /GENE=glgB /DEF=1,4-alpha-glucan branching enzyme /FUNCTION=Energy metabolism
21.9	0.2897	A	A	PA2165 /DEF=probable glycogen synthase /FUNCTION=Energy metabolism
0.5	0.4840	A	P	PA2250 /GENE=lpdV /DEF=lipoamide dehydrogenase-Val /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism
0.2	0.0178	A	P	PA2266 /DEF=probable cytochrome c precursor /FUNCTION=Carbon compound catabolism; Energy metabolism
0.1	0.0274	A	P	PA2290 /GENE=gcd /DEF=glucose dehydrogenase /FUNCTION=Carbon compound catabolism; Energy metabolism
7.4	0.0918	A	P	PA2297 /DEF=probable ferredoxin /FUNCTION=Energy metabolism
0.8	0.6013	A	P	PA2321 /DEF=gluconokinase /FUNCTION=Carbon compound catabolism; Energy metabolism

Table A.3. (contd.)

3.1	0.9490	A	P	PA2382 /GENE=lldA /DEF=L-lactate dehydrogenase /FUNCTION=Energy metabolism				
0.9	0.9216	A	P	PA2482 /DEF=probable cytochrome c /FUNCTION=Energy metabolism				
4.1	0.1929	A	P	PA2516 /GENE=xylZ /DEF=toluate 1,2-dioxygenase electron transfer component /FUNCTION=Energy metabolism; Carbon compound catabolism				
0.0	0.3024	A	P	PA2623 /GENE=icd /DEF=isocitrate dehydrogenase /FUNCTION=Amino acid biosynthesis and metabolism; Carbon compound catabolism; Energy m				
0.7	0.5642	P	P	PA2624 /GENE=idh /DEF=isocitrate dehydrogenase /FUNCTION=Energy metabolism				
1.2	0.9506	P	P	PA2644 /GENE=nuoI /DEF=NADH Dehydrogenase I chain I /FUNCTION=Energy metabolism				
0.2	0.3442	A	P	PA2645 /GENE=nuoJ /DEF=NADH dehydrogenase I chain J /FUNCTION=Energy metabolism				
0.5	0.4635	A	P	PA2646 /GENE=nuoK /DEF=NADH dehydrogenase I chain K /FUNCTION=Energy metabolism				
0.5	0.4054	A	P	PA2647 /GENE=nuoL /DEF=NADH dehydrogenase I chain L /FUNCTION=Energy metabolism				
1.9	0.6489	P	P	PA2648 /GENE=nuoM /DEF=NADH dehydrogenase I chain M /FUNCTION=Energy metabolism				
0.9	0.6474	A	P	PA2649 /GENE=nuoN /DEF=NADH dehydrogenase I chain N /FUNCTION=Energy metabolism				
3.5	0.7247	A	P	PA2664 /GENE=fhp /DEF=flavohemoprotein /FUNCTION=Energy metabolism				
27.3	0.2636	A	A	PA2680 /DEF=probable quinone oxidoreductase /FUNCTION=Energy metabolism				
13.5	0.0383	A	A	PA2694 /DEF=probable thioredoxin /FUNCTION=Energy metabolism				
21.1	0.1344	A	A	PA2714 /DEF=probable molybdopterin oxidoreductase /FUNCTION=Energy metabolism				
2.6	0.2698	A	A	PA2715 /DEF=probable ferredoxin /FUNCTION=Energy metabolism				
3.0	0.4139	A	A	PA2716 /DEF=probable FMN oxidoreductase /FUNCTION=Energy metabolism				
0.5	0.3527	A	P	PA2796 /GENE=tal /DEF=transaldolase /FUNCTION=Carbon compound catabolism; Energy metabolism				
0.1	0.3314	A	P	PA2951 /GENE=etfA /DEF=electron transfer flavoprotein alpha-subunit /FUNCTION=Energy metabolism				
0.6	0.5247	A	P	PA2952 /GENE=etfB /DEF=electron transfer flavoprotein beta-subunit /FUNCTION=Energy metabolism				
3.2	0.2168	P	P	PA2953 /DEF=electron transfer flavoprotein-ubiquinone oxidoreductase /FUNCTION=Energy metabolism				
9.9	0.1014	A	P	PA2994 /GENE=nqrF /DEF=Na+-translocating NADH:quinone oxidoreductase, subunit Nqr6 /FUNCTION=Energy metabolism				
12.2	0.1597	P	P	PA2995 /GENE=nqrE /DEF=Na+-translocating NADH:quinone oxidoreductase subunit Nqr5 /FUNCTION=Energy metabolism				
1.3	0.9225	A	P	PA2996 /GENE=nqrD /DEF=Na+-translocating NADH:uniquinone oxidoreductase subunit Nqr4 /FUNCTION=Energy metabolism				

Table A.3. (contd.)

3.5	0.0774	P	P	PA3194 /GENE=edd /DEF=phosphogluconate dehydratase /FUNCTION=Carbon compound catabolism; Energy metabolism
0.4	0.1098	A	P	PA3195 /GENE=gapA /DEF=glyceraldehyde 3-phosphate dehydrogenase /FUNCTION=Energy metabolism; Carbon compound catabolism
0.2	0.4085	A	P	PA3391 /GENE=nosR /DEF=regulatory protein NosR /FUNCTION=Energy metabolism; Membrane proteins
0.2	0.4325	A	P	PA3392 /GENE=nosZ /DEF=nitrous-oxide reductase precursor /FUNCTION=Energy metabolism
3.4	0.9859	P	P	PA3393 /GENE=nosD /DEF=NosD protein /FUNCTION=Energy metabolism
5.6	0.6661	A	P	PA3394 /GENE=nosF /DEF=NosF protein /FUNCTION=Energy metabolism; Transport of small molecules
9.4	0.6502	A	P	PA3395 /GENE=nosY /DEF=NosY protein /FUNCTION=Energy metabolism; Membrane proteins
1.6	0.8120	A	P	PA3396 /GENE=nosL /DEF=NosL protein /FUNCTION=Energy metabolism
2.0	0.9640	A	P	PA3397 /GENE=fpr /DEF=ferredoxin--NADP+ reductase /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabolism
1.5	0.6281	A	P	PA3415 /DEF=probable dihydrolipoamide acetyltransferase /FUNCTION=Energy metabolism
1.2	0.5188	A	P	PA3416 /DEF=probable pyruvate dehydrogenase E1 component, beta chain /FUNCTION=Energy metabolism
1.9	0.7563	P	P	PA3417 /DEF=probable pyruvate dehydrogenase E1 component, alpha subunit /FUNCTION=Energy metabolism
1.4	0.7042	A	P	PA3452 /GENE=mqoA /DEF=malate:quinone oxidoreductase /FUNCTION=Central intermediary metabolism; Energy metabolism
2.3	0.8088	A	P	PA3490 /DEF=probable ferredoxin /FUNCTION=Energy metabolism
3.1	0.2612	P	P	PA3491 /DEF=probable ferredoxin /FUNCTION=Energy metabolism; Membrane proteins
0.2	0.1469	P	P	PA3584 /GENE=glpD /DEF=glycerol-3-phosphate dehydrogenase /FUNCTION=Central intermediary metabolism; Energy metabolism
0.7	0.5614	A	P	PA3621 /GENE=fdxA /DEF=ferredoxin I /FUNCTION=Energy metabolism
2.3	0.4316	P	P	PA3635 /GENE=eno /DEF=enolase /FUNCTION=Energy metabolism; Translation, post-translational modification, degradation; Carbon compound ca
0.2	0.3247	A	P	PA3636 /GENE=kdsA /DEF=2-dehydro-3-deoxyphosphooctonate aldolase /FUNCTION=Energy metabolism; Translation,
3.7	0.1663	A	P	PA3687 /GENE=ppc /DEF=phosphoenolpyruvate carboxylase /FUNCTION=Central intermediary metabolism; Energy metabolism
0.7	0.8558	A	P	PA3809 /GENE=fdx2 /DEF=ferredoxin [2Fe-2S] /FUNCTION=Energy metabolism
0.4	0.4475	A	P	PA3872 /GENE=narI /DEF=respiratory nitrate reductase gamma chain /FUNCTION=Energy metabolism
5.6	0.5559	P	P	PA3873 /GENE=narJ /DEF=respiratory nitrate reductase delta chain /FUNCTION=Energy metabolism
0.5	0.4299	A	P	PA3874 /GENE=narH /DEF=respiratory nitrate reductase beta chain /FUNCTION=Energy metabolism
3.0	0.5115	A	P	PA3875 /GENE=narG /DEF=respiratory nitrate reductase alpha chain /FUNCTION=Energy metabolism
1.5	0.5450	A	P	PA3878 /GENE=narX /DEF=two-component sensor NarX /FUNCTION=Energy metabolism; Two-component regulatory systems

Table A.3. (contd.)

2.7	0.3248	A	P	PA3879 /GENE=narL /DEF=two-component response regulator NarL /FUNCTION=Energy metabolism; Two-component regulatory systems
1.8	0.2867	A	P	PA3929 /GENE=cioB /DEF=cyanide insensitive terminal oxidase /FUNCTION=Energy metabolism
0.4	0.2256	A	P	PA3930 /GENE=cioA /DEF=cyanide insensitive terminal oxidase /FUNCTION=Energy metabolism
1.6	0.4608	A	P	PA4061 /DEF=probable thioredoxin /FUNCTION=Energy metabolism
8.4	0.0160	P	A	PA4470 /GENE=fumC1 /DEF=fumarate hydratase /FUNCTION=Energy metabolism
0.4	0.2699	A	P	PA4538 /GENE=ndh /DEF=NADH dehydrogenase /FUNCTION=Energy metabolism
0.5	0.3974	A	P	PA4569 /GENE=ispB /DEF=octaprenyl-diphosphate synthase /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabol
1.0	0.5968	A	P	PA4571 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
0.2	0.4326	A	P	PA4587 /GENE=ccpR /DEF=cytochrome c551 peroxidase precursor /FUNCTION=Energy metabolism
0.4	0.6535	A	P	PA4640 /GENE=mqoB /DEF=malate:quinone oxidoreductase /FUNCTION=Central intermediary metabolism; Energy metabolism
8.8	0.0026	P	P	PA4732 /GENE=pgi /DEF=glucose-6-phosphate isomerase /FUNCTION=Carbon compound catabolism; Energy metabolism
0.2	0.2993	A	P	PA4748 /GENE=tpiA /DEF=triosephosphate isomerase /FUNCTION=Central intermediary metabolism; Energy metabolism
3.6	0.2175	A	P	PA4771 /GENE=lldD /DEF=L-lactate dehydrogenase /FUNCTION=Energy metabolism
2.4	0.9719	A	P	PA4772 /DEF=probable ferredoxin /FUNCTION=Energy metabolism
0.6	0.5168	A	P	PA4809 /GENE=fdhE /DEF=FdhE protein /FUNCTION=Energy metabolism
0.2	0.1588	A	P	PA4810 /GENE=fdnI /DEF=nitrate-inducible formate dehydrogenase, gamma subunit /FUNCTION=Energy metabolism
0.1	0.0062	A	P	PA4811 /GENE=fdnH /DEF=nitrate-inducible formate dehydrogenase, beta subunit /FUNCTION=Energy metabolism
0.6	0.3379	A	P	PA4812 /GENE=fdnG /DEF=formate dehydrogenase-O, major subunit /FUNCTION=Energy metabolism
3.8	0.2468	P	P	PA4829 /GENE=lpd3 /DEF=dihydrolipoamide dehydrogenase 3 /FUNCTION=Energy metabolism
0.0	0.2812	A	P	PA4922 /GENE=azu /DEF=azurin precursor /FUNCTION=Energy metabolism
10.1	0.0409	A	A	PA4975 /DEF=NAD(P)H quinone oxidoreductase /FUNCTION=Energy metabolism
1.5	0.9112	P	P	PA5015 /GENE=aceE /DEF=pyruvate dehydrogenase /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism
0.9	0.6545	P	P	PA5016 /GENE=aceF /DEF=dihydrolipoamide acetyltransferase /FUNCTION=Carbon compound catabolism; Energy metabolism
1.0	0.7581	P	P	PA5063 /GENE=ubiE /DEF=ubiquinone biosynthesis methyltransferase UbiE /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers;
0.2	0.3478	A	P	PA5129 /GENE=grx /DEF=glutaredoxin /FUNCTION=Energy metabolism; Nucleotide biosynthesis and metabolism
2.4	0.2533	P	P	PA5192 /GENE=pckA /DEF=phosphoenolpyruvate carboxykinase /FUNCTION=Carbon compound catabolism; Energy metabolism

Table A.3. (contd.)

2.0	0.2735	A	P	PA5223 /GENE=ubiH /DEF=ubiH protein /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabolism				
0.3	0.4713	A	P	PA540 /GENE=trxA /DEF=thioredoxin /FUNCTION=Nucleotide biosynthesis and metabolism; Translation, post-translational modification, degradation;				
7.9	0.0394	A	A	PA5297 /GENE=poxB /DEF=pyruvate dehydrogenase (cytochrome) /FUNCTION=Central intermediary metabolism; Energy metabolism				
1.4	0.5677	P	P	PA5300 /GENE=cycB /DEF=cytochrome c5 /FUNCTION=Energy metabolism				
0.7	0.8094	A	P	PA5304 /GENE=dadA /DEF=D-amino acid dehydrogenase, small subunit /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism				
0.1	0.3071	A	P	PA5491 /DEF=probable cytochrome /FUNCTION=Energy metabolism				
1.2	0.6740	P	P	PA5553 /GENE=atpC /DEF=ATP synthase epsilon chain /FUNCTION=Energy metabolism				
2.8	0.8189	P	P	PA5554 /GENE=atpD /DEF=ATP synthase beta chain /FUNCTION=Energy metabolism				
0.0	0.4159	A	P	PA5555 /GENE=atpG /DEF=ATP synthase gamma chain /FUNCTION=Energy metabolism				
1.5	0.6360	P	P	PA5556 /GENE=atpA /DEF=ATP synthase alpha chain /FUNCTION=Energy metabolism				
0.1	0.4064	A	P	PA5557 /GENE=atpH /DEF=ATP synthase delta chain /FUNCTION=Energy metabolism				
0.0	0.3952	A	P	PA5558 /GENE=atpF /DEF=ATP synthase B chain /FUNCTION=Energy metabolism				
0.0	0.3936	A	P	PA5559 /GENE=atpE /DEF=atp synthase C chain /FUNCTION=Energy metabolism				
0.6	0.4947	P	P	PA5560 /GENE=atpB /DEF=ATP synthase A chain /FUNCTION=Energy metabolism				
0.3	0.2705	A	P	PA5561 /GENE=atpI /DEF=ATP synthase protein I /FUNCTION=Energy metabolism; Membrane proteins				

Table A.4. Sub-groups of energy-metabolism genes

ATP synthases													
		Call Detection											
Fold change	**p-value**	**Av**	**Pae**		**Gene description**								
2.0	0.6518	A	P	PA0649 /GENE=trpG /DEF=anthranilate synthase component II /FUNCTION=Amino acid biosynthesis and metabolism;									
2.0	0.2721	A	P	PA1104 /GENE=flil /DEF=flagellum-specific ATP synthase FliI /FUNCTION=Motility & Attachment; Energy metabolism									
0.2	0.3610	A	P	PA1580 /GENE=gltA /DEF=citrate synthase /FUNCTION=Energy metabolism									
0.2	0.4043	A	P	PA1770 /GENE=ppsA /DEF=phosphoenolpyruvate synthase /FUNCTION=Carbon compound catabolism;									
21.9	0.2897	A	A	PA2165 /DEF=probable glycogen synthase /FUNCTION=Energy metabolism									
0.5	0.3974	A	P	PA4569 /GENE=ispB /DEF=octaprenyl-diphosphate synthase /FUNCTION=Biosynthesis of cofactors, prosthetic groups and									
1.2	0.6740	P	P	PA5553 /GENE=atpC /DEF=ATP synthase epsilon chain /FUNCTION=Energy metabolism									
2.8	0.8189	P	P	PA5554 /GENE=atpD /DEF=ATP synthase beta chain /FUNCTION=Energy metabolism									
0.0	0.4159	A	P	PA5555 /GENE=atpG /DEF=ATP synthase gamma chain /FUNCTION=Energy metabolism									
1.5	0.6360	P	P	PA5556 /GENE=atpA /DEF=ATP synthase alpha chain /FUNCTION=Energy metabolism									
0.1	0.4064	A	P	PA5557 /GENE=atpH /DEF=ATP synthase delta chain /FUNCTION=Energy metabolism									
0.0	0.3952	A	P	PA5558 /GENE=atpF /DEF=ATP synthase B chain /FUNCTION=Energy metabolism									
0.0	0.3936	A	P	PA5559 /GENE=atpE /DEF=atp synthase C chain /FUNCTION=Energy metabolism									
0.6	0.4947	P	P	PA5560 /GENE=atpB /DEF=ATP synthase A chain /FUNCTION=Energy metabolism									
0.3	0.2705	A	P	PA5561 /GENE=atpI /DEF=ATP synthase protein I /FUNCTION=Energy metabolism; Membrane proteins									

Table A.4. (contd.)

Carbon compound metabolism													
Fold change	**p-value**	**Av**	**Pae**		**Gene description**								
9.9	0.2432	A	P	PA0552 /GENE=pgk /DEF=phosphoglycerate kinase /FUNCTION=Carbon compound catabolism; Energy metabolism									
0.8	0.9822	A	P	PA0854 /GENE=fumC2 /DEF=fumarate hydratase /FUNCTION=Carbon compound catabolism; Energy metabolism									
3.4	0.4269	A	P	PA0927 /GENE=ldhA /DEF=D-lactate dehydrogenase (fermentative) /FUNCTION=Carbon compound catabolism; Central intermediary metabolism; Ene									
2.0	0.5171	A	A	PA1498 /GENE=pykF /DEF=pyruvate kinase I /FUNCTION=Carbon compound catabolism; Energy metabolism									
0.2	0.4043	A	P	PA1770 /GENE=ppsA /DEF=phosphoenolpyruvate synthase /FUNCTION=Carbon compound catabolism; Central intermediary metabolism; Energy me									
1.6	0.8346	A	P	PA1931 /DEF=probable ferredoxin /FUNCTION=Carbon compound catabolism; Energy metabolism									
2.2	0.3797	P	P	PA3183 /GENE=zwf /DEF=glucose-6-phosphate 1-dehydrogenase /FUNCTION=Carbon compound catabolism; Energy metabolism									
4.1	0.4741	A	P	PA3193 /GENE=glk /DEF=glucokinase /FUNCTION=Carbon compound catabolism; Energy metabolism									
3.5	0.0774	P	P	PA3194 /GENE=edd /DEF=phosphogluconate dehydratase /FUNCTION=Carbon compound catabolism; Energy metabolism									
0.4	0.1098	A	P	PA3195 /GENE=gapA /DEF=glyceraldehyde 3-phosphate dehydrogenase /FUNCTION=Energy metabolism; Carbon compound catabolism									
2.3	0.4316	P	P	PA3635 /GENE=eno /DEF=enolase /FUNCTION=Energy metabolism; Translation, post-translational modification, degradation; Carbon compound cat									
0.2	0.3247	A	P	PA3636 /GENE=kdsA /DEF=2-dehydro-3-deoxyphosphooctonate aldolase /FUNCTION=Energy metabolism; Translation, post-translational									
1.6	0.9686	A	P	PA4329 /GENE=pykA /DEF=pyruvate kinase II /FUNCTION=Energy metabolism; Carbon compound catabolism									
8.8	0.0026	P	P	PA4732 /GENE=pgi /DEF=glucose-6-phosphate isomerase /FUNCTION=Carbon compound catabolism; Energy metabolism									
0.9	0.6545	P	P	PA5016 /GENE=aceF /DEF=dihydrolipoamide acetyltransferase /FUNCTION=Carbon compound catabolism; Energy metabolism									
2.4	0.2533	P	P	PA5192 /GENE=pckA /DEF=phosphoenolpyruvate carboxykinase /FUNCTION=Carbon compound catabolism; Energy metabolism									
0.2	0.4071	A	P	PA5332 /GENE=crc /DEF=catabolite repression control protein /FUNCTION=Carbon compound catabolism; Energy metabolism									

Table A.4. (contd.)

Cofactor biosynthesis													
Fold change	**p-value**	**Av**	**Pae**		**Gene description**								
2.3	0.9963	P	P	PA0509 /GENE=nirN /DEF=probable c-type cytochrome /FUNCTION=Energy metabolism; Biosynthesis of cofactors, prosthetic groups and carriers									
0.2	0.4219	A	P	PA0510 /DEF=probable uroporphyrin-III c-methyltransferase /FUNCTION=Energy metabolism; Biosynthesis of cofactors, prosthetic groups and carr									
2.5	0.9376	A	P	PA0511 /GENE=nirJ /DEF=heme d1 biosynthesis protein NirJ /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabol									
0.2	0.4257	A	P	PA0512 /DEF=conserved hypothetical protein /FUNCTION=Energy metabolism; Hypothetical, unclassified, unknown; Biosynthesis of cofactors, prosth									
2.1	0.9481	A	P	PA0513 /DEF=probable transcriptional regulator /FUNCTION=Energy metabolism; Transcriptional regulators; Biosynthesis of cofactors, prosthetic gr									
0.3	0.4466	A	P	PA0514 /GENE=nirL /DEF=heme d1 biosynthesis protein NirL /FUNCTION=Energy metabolism; Hypothetical, unclassified, unknown;									
2.0	0.2735	A	P	PA5223 /GENE=ubiH /DEF=ubiH protein /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; Energy metabolism									
5.7	0.0162	P	P	PA5358 /GENE=ubiA /DEF=4-hydroxybenzoate-octaprenyl transferase /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers									
Cytochromes													
Fold change	**p-value**	**Av**	**Pae**		**Gene description**								
0.4	0.4580	P	P	PA0105 /GENE=coxB /DEF=cytochrome c oxidase, subunit II /FUNCTION=Energy metabolism									
3.6	0.5614	A	P	PA0106 /GENE=coxA /DEF=cytochrome c oxidase, subunit I /FUNCTION=Energy metabolism									
0.4	0.4450	P	P	PA0108 /GENE=coIII /DEF=cytochrome c oxidase, subunit III /FUNCTION=Energy metabolism									
1.6	0.6586	P	P	PA0113 /DEF=probable cytochrome c oxidase assembly factor /FUNCTION=Energy metabolism									
2.3	0.9963	P	P	PA0509 /GENE=nirN /DEF=probable c-type cytochrome /FUNCTION=Energy metabolism; Biosynthesis of cofactors, prosthetic gr									
0.1	0.4249	A	P	PA0517 /GENE=nirC /DEF=probable c-type cytochrome precursor /FUNCTION=Biosynthesis of cofactors, prosthetic groups and c									
0.4	0.4390	A	P	PA0518 /GENE=nirM /DEF=cytochrome c-551 precursor /FUNCTION=Biosynthesis of cofactors, prosthetic groups and carriers; E									
3.1	0.3772	A	P	PA0521 /DEF=probable cytochrome c oxidase subunit /FUNCTION=Energy metabolism									
0.3	0.0470	A	P	PA0918 /DEF=cytochrome b561 /FUNCTION=Energy metabolism									
0.3	0.3847	A	P	PA1172 /GENE=napC /DEF=cytochrome c-type protein NapC /FUNCTION=Energy metabolism									
0.6	0.5562	P	P	PA1173 /GENE=napB /DEF=cytochrome c-type protein NapB precursor /FUNCTION=Energy metabolism									
7.5	0.3583	A	A	PA1317 /GENE=cyoA /DEF=cytochrome o ubiquinol oxidase subunit II /FUNCTION=Energy metabolism									
6.2	0.0444	A	A	PA1318 /GENE=cyoB /DEF=cytochrome o ubiquinol oxidase subunit I /FUNCTION=Energy metabolism									

Table A.4. (contd.)

1.3	0.7097	P	P	PA1482 /GENE=ccmH /DEF=cytochrome C-type biogenesis protein CcmH /FUNCTION=Energy metabolism
0.3	0.4512	A	P	PA1483 /GENE=cycH /DEF=cytochrome c-type biogenesis protein /FUNCTION=Energy metabolism
0.6	0.6999	A	P	PA1552 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
0.2	0.3484	A	P	PA1553 /DEF=probable cytochrome c oxidase subunit /FUNCTION=Energy metabolism
0.3	0.3687	A	P	PA1554 /DEF=probable cytochrome oxidase subunit (cbb3-type) /FUNCTION=Energy metabolism
0.2	0.4327	A	P	PA1555 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
0.3	0.4335	A	P	PA1556 /DEF=probable cytochrome c oxidase subunit /FUNCTION=Energy metabolism
0.2	0.4316	A	P	PA1557 /DEF=probable cytochrome oxidase subunit (cbb3-type) /FUNCTION=Energy metabolism
2.5	0.2962	P	A	PA1600 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
1.1	0.9484	A	P	PA1983 /GENE=exaB /DEF=cytochrome c550 /FUNCTION=Carbon compound catabolism; Energy metabolism
0.2	0.0178	A	P	PA2266 /DEF=probable cytochrome c precursor /FUNCTION=Carbon compound catabolism; Energy metabolism
0.9	0.9216	A	P	PA2482 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
2.3	0.9927	P	P	PA3032 /DEF=cytochrome c /FUNCTION=Energy metabolism
1.0	0.8734	P	P	PA4133 /DEF=cytochrome c oxidase subunit (cbb3-type) /FUNCTION=Energy metabolism
0.2	0.4381	A	P	PA4429 /DEF=probable cytochrome c1 precursor /FUNCTION=Energy metabolism
0.1	0.4296	A	P	PA4430 /DEF=probable cytochrome b /FUNCTION=Energy metabolism
1.0	0.5968	A	P	PA4571 /DEF=probable cytochrome c /FUNCTION=Energy metabolism
0.2	0.4326	A	P	PA4587 /GENE=ccpR /DEF=cytochrome c551 peroxidase precursor /FUNCTION=Energy metabolism
7.9	0.0394	A	A	PA5297 /GENE=poxB /DEF=pyruvate dehydrogenase (cytochrome) /FUNCTION=Central intermediary metabolism; Energy metabo
1.4	0.5677	P	P	PA5300 /GENE=cycB /DEF=cytochrome c5 /FUNCTION=Energy metabolism
3.6	0.0568	A	A	PA5328 /DEF=probable cytochrome c(mono-heme type) /FUNCTION=Energy metabolism

Table A.4. (contd.)

Quinones													
Fold change	**p-value**	**Av**	**Pae**		**Gene description**								
1.8	0.5057	A	P	PA0023 /GENE=qor /DEF=quinone oxidoreductase /FUNCTION=Energy metabolism									
4.1	0.2119	A	A	PA1883 /DEF=probable NADH-ubiquinone/plastoquinone oxidoreductase /FUNCTION=Energy metabolism									
27.3	0.2636	A	A	PA2680 /DEF=probable quinone oxidoreductase /FUNCTION=Energy metabolism									
3.2	0.2168	P	P	PA2953 /DEF=electron transfer flavoprotein-ubiquinone oxidoreductase /FUNCTION=Energy metabolism									
0.2	0.4065	A	P	PA2999 /GENE=nqrA /DEF=Na+-translocating NADH:ubiquinone oxidoreductase subunit Nrq1 /FUNCTION=Energy metabolism									
1.0	0.7024	A	P	PA3171 /GENE=ubiG /DEF=3-demethylubiquinone-9 3-methyltransferase /FUNCTION=Biosynthesis of cofactors, prosthetic									
1.4	0.7042	A	P	PA3452 /GENE=mqoA /DEF=malate:quinone oxidoreductase /FUNCTION=Central intermediary metabolism; Energy metabolism									
0.4	0.6535	A	P	PA4640 /GENE=mqoB /DEF=malate:quinone oxidoreductase /FUNCTION=Central intermediary metabolism; Energy metabolism									
10.1	0.0409	A	A	PA4975 /DEF=NAD(P)H quinone oxidoreductase /FUNCTION=Energy metabolism									
1.0	0.7581	P	P	PA5063 /GENE=ubiE /DEF=ubiquinone biosynthesis methyltransferase UbiE /FUNCTION=Biosynthesis of cofactors, prosthetic									
Amino acid biosynthesis													
Fold change	**p-value**	**Av**	**Pae**		**Gene description**								
1.8	0.0737	A	P	PA0609 /GENE=trpE /DEF=anthranilate synthetase component I /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism									
2.0	0.6518	A	P	PA0649 /GENE=trpG /DEF=anthranilate synthase component II /FUNCTION=Amino acid biosynthesis and metabolism; Biosynthesis of cofactors, pros									
0.2	0.3902	A	P	PA1585 /GENE=sucA /DEF=2-oxoglutarate dehydrogenase (E1 subunit) /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism									
0.1	0.3547	A	P	PA1587 /GENE=lpdG /DEF=lipoamide dehydrogenase-glc /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism									
0.5	0.4840	A	P	PA2250 /GENE=lpdV /DEF=lipoamide dehydrogenase-Val /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism									
0.0	0.3024	A	P	PA2623 /GENE=icd /DEF=isocitrate dehydrogenase /FUNCTION=Amino acid biosynthesis and metabolism; Carbon compound catabolism; Energy									
1.5	0.9112	P	P	PA5015 /GENE=aceE /DEF=pyruvate dehydrogenase /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism									
0.7	0.8094	A	P	PA5304 /GENE=dadA /DEF=D-amino acid dehydrogenase, small subunit /FUNCTION=Amino acid biosynthesis and metabolism; Energy metabolism									

Table A.5. Energy metabolism genes that show significantly higher expression in *A. vinelandii* as compared to *P. aeruginosa*

Gene	SI Azo1	SI Azo2	SI Pae1	SI Pae2	Fold chang	p-value	Description
PA0330_rpiA_at	112.8	130.5	42.2	13.7	5.0	0.0306	PA0330 /GENE=rpiA /DEF=ribose 5-phosphate isomerase /FUNCTION=Energy metabolism
PA1224_at	32.9	43.1	4.6	1.8	13.1	0.0223	PA1224 /DEF=probable NAD(P)H dehydrogenase /FUNCTION=Putative enzymes
PA1318_cyoB_at	30.8	37.8	12	2.5	6.2	0.0444	PA1318 /GENE=cyoB /DEF=cytochrome o ubiquinol oxidase subunit I /FUNCTION=Energy metabolism
PA2153_glgB_at	38.4	40.5	6.2	3.8	8.1	0.0021	PA2153 /GENE=glgB /DEF=1,4-alpha-glucan branching enzyme /FUNCTION=Energy metabolism
PA2357_msuE_at	20.7	17.9	6	0.4	12.4	0.0358	PA2357 /GENE=msuE /DEF=NADH-dependent FMN reductase MsuE /FUNCTION=Central intermediary metabolis
PA2694_at	30.1	42.8	1.3	5.4	13.5	0.0383	PA2694 /DEF=probable thioredoxin /FUNCTION=Energy metabolism
PA4331_at	20.8	27.2	7.5	8.7	2.9	0.0395	PA4331 /DEF=probable ferredoxin reductase /FUNCTION=Energy metabolism
PA4470_fumC1_at	47.8	38.1	5.1	5.1	8.4	0.0160	PA4470 /GENE=fumC1 /DEF=fumarate hydratase /FUNCTION=Energy metabolism
PA4732_pgi_at	254.4	242.6	39.3	20.3	8.8	0.0026	PA4732 /GENE=pgi /DEF=glucose-6-phosphate isomerase /FUNCTION=Carbon compound cataboli
PA4975_at	21.8	30.7	5	1.3	10.1	0.0409	PA4975 /DEF=NAD(P)H quinone oxidoreductase /FUNCTION=Energy metabolism
PA5297_poxB_at	40.2	57.1	8.2	4.5	7.9	0.0394	PA5297 /GENE=poxB /DEF=pyruvate dehydrogenase (cytochrome) /FUNCTION=Central intermediary metabolism;
PA5358_ubiA_at	51.7	60.6	14.4	6.7	5.7	0.0162	PA535/GENE=ubiA /DEF=4-hydroxybenzoate-octaprenyl transferase /FUNCTION=Biosynthesis of cofactors, pro
*SI = Signal Intensity							
Azo1, Azo 2= Microarray experiments utilizing cDNA from *Azotobacter vinelandii*							
Pae1, Pae 2= Microarray experiments utilizing cDNA from *Pseduomonas aeruginosa*							

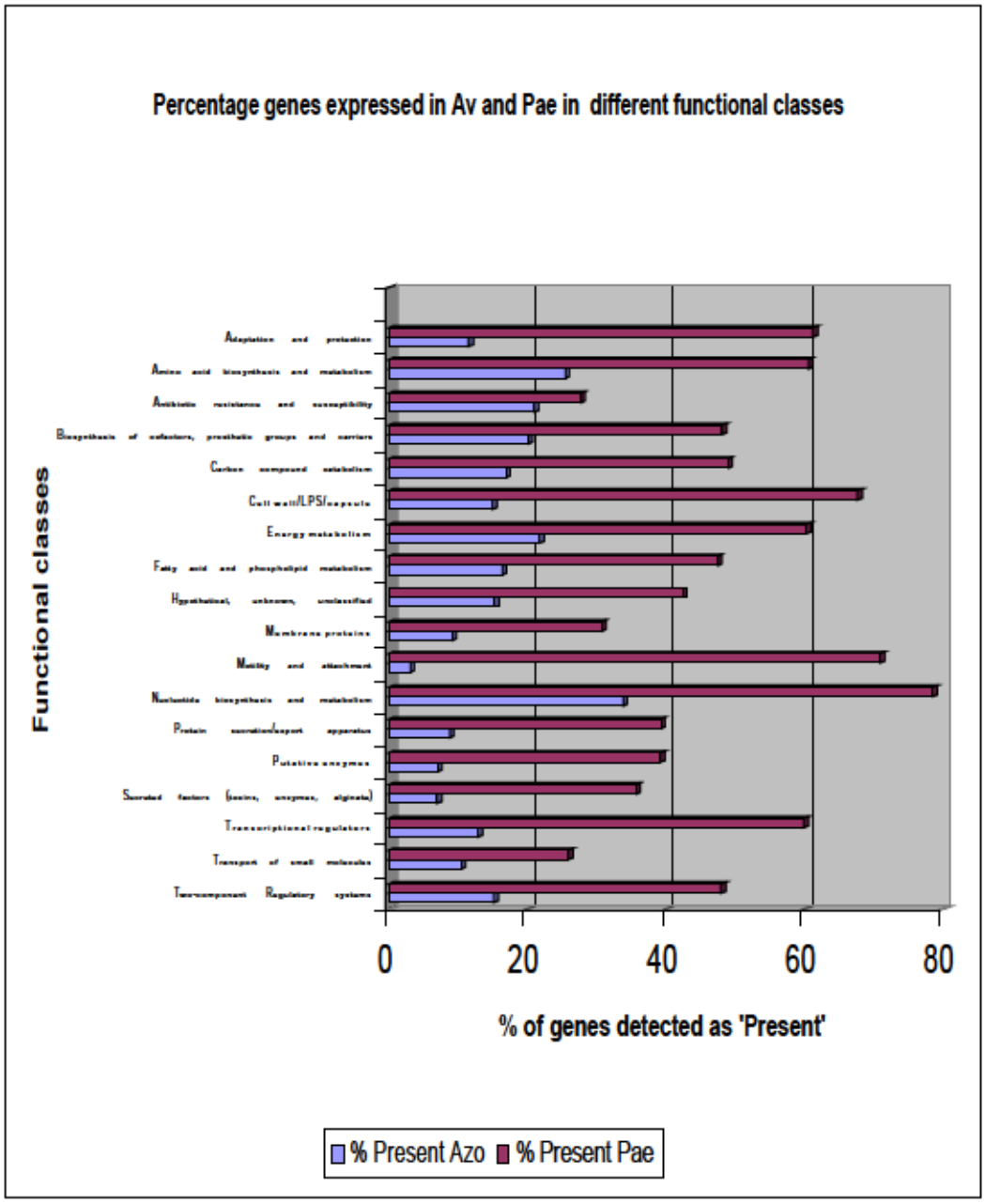


Fig. A.4. Percentage genes with 'Present' call in *A. vinelandii* and *P. aeruginosa* sorted according to their functional classes.

Table A.6. Detailed description of highly expressed energy metabolism genes in *A. vinelandii* in comparison to *P. aeruginosa*

***P. aeruginosa* gene**	**Gene title**	**Fold change**	**Pathway**	**Specific role of gene product**
PA0330 *rpiA*	Ribose 5-phosphate isomerase	5.0	Pentose-phosphate pathway	Catalyzes the conversion of Ribose 5-phosphate to Ribulose 5-phosphate
PA4470 *fumC1*	Fumarate hydratase	8.4	Citrate cycle (TCA cycle)	Catalyzes the conversion of malate to fumarate
PA4732 *pgi*	Glucose 6-phosphate isomerase	8.8	Glycolysis / Gluconeogenesis, Pentose phosphate pathway, Starch and sucrose metabolism	Catalyzes the conversion of D-glucose 6-phosphate to D-fructose 6-phosphate
PA5358 *ubiA*	4-hydroxybenzoate-octaprenyl transferase	5.7	Ubiquinone biosynthesis	catalyzes the prenylation of 4-hydroxybenzoic acid
PA1318 *cyoB*	cytochrome o ubiquinol oxidase	6.2	Energy metabolism	Functions as terminal oxidase in electron transport chain
PA2153 *glgB*	1,4-alpha-glucan branching enzyme	8.1	Starch and sucrose metabolism	Converts amylose into amylopectin.
PA2694	Probable thioredoxin	13.5	Energy metabolism	Serves as a general protein disulphide oxidoreductase
PA4331	Probable ferredoxin reductase	2.9	Energy metabolism	Iron reduction
PA4975	NAD(P)H quinone oxidoreductase	10.1	Energy metabolism	catalyzes the two-electron reduction of quinoid compounds into hydroquinones
PA5297 *poxB*	Pyruvate dehydrogenase (cytochrome)	7.9	Energy metabolism	Catalyzes the reaction: pyruvate + ferricytochrome b1 + H2O = acetate + CO2 + ferrocytochrome b1
PA1224	probable NAD(P)H dehydrogenase	13.1	Energy metabolism	NAD(P)H + H^+ + acceptor = $NADP^+$ + reduced acceptor
PA2357 *msuE*	NADH-dependent FMN reductase	12.4	Central intermediary metabolism, Carbon compound catabolism	Catalysis of the reaction: FMN + NADH = FMNH2 + NAD^+

Previous research has shown that *Azotobacters* have a very high rate of respiration and this oxygen scavenging activity is largely responsible for the protection of its oxygen sensitive nitrogenase enzyme (Drozd and Postgate, 1970). This led to the hypothesis that there must be more efficient energy metabolism in *A. vinelandii* as compared to the closely related non-nitrogen fixing *P. aeruginosa* bacterium even in normal growth conditions. We therefore used the *P. aeruginosa* PAO1 DNA chip for comparison of energy metabolism genes in *A. vinelandii* and *P. aeruginosa*.

Our work has provided initial data on the differential gene expression between *P. aeruginosa* and *A. vinelandii* when grown in the same media. We have listed energy metabolism related genes showing a significantly higher gene expression in *A. vinelandii* as compared to *P. aeruginosa*. Further validation of the differential expression of these set of genes can be obtained by performing real time RT PCR. A detailed study on the respiratory chain of *A. vinelandii* will prove to be useful in providing an overall view of the role of the differentially expressed genes detected in our investigation. Also, future microarray studies involving expression of genes in *A. vinelandii* grown in BN- media (that allows nitrogen fixation to occur because of unavailability of fixed nitrogen in the media) will enable us to compare those energy metabolism genes that are being expressed highly in response to nitrogen fixing activity versus those that we have detected in this study (under nitrogen non-fixing conditions). Finally, this work is important since we have shown here for the first time that based on the relatedness of the *A. vinelandii* and *P. aeruginosa* genomes, the *P. aeruginosa* DNA chip could be successfully used for investigating the global gene expression of *A. vinelandii*.

References

Bergman, B., Gallon, J. R., Rai, A. N. and Stal, L. J. 1997. N2 fixation by non-heterocystous cyanobacteria. FEMS Microbiol. Rev. 19: 139-185

Drozd, J. and Postgate, J. R. 1970. Effect of oxygen on acetylene reduction, cytochrome content and respiratory activity of *Azotobacter chroococcum*. J. Gen. Microbiol. 63: 63-73

Gallon, J. R. 1992. Reconciling the incompatible: N_2 fixation and O_2. New Phytol. 122: 571-609

Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc J. J. and Rees, D.C. 1992. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii.* Science. 257(5077): 1653-9

Hill, S. Physiology of nitrogen fixation in free-living heterotrophs, 1992, In: Biological Nitrogen Fixation (Stacey, G., Burris, R. H. and Evans, H. J., Eds.) pp 87-133, Chapman and Hall, London.

Kim, J. S. and Rees, D. C. 1992. Crystallographic structure and functional implications of the nitrogenase molybdenum iron protein from *Azotobacter vinelandii*. Nature, 360(6404): 553-560.

Linkerhägner K and Oelze J. 1995. Cellular ATP level and nitrogenase switch off upon oxygen stress in chemostat cultures of *Azotobacter vinelandii*. J. Bacteriol. 177: 5289–5293.

Liu, J. K., Lee, F. T., Lin, C. S., Yao, X. T., Davenport, J. W. and Wong, T. Y. 1995. Alternative function of the electron transport system in *Azotobacter vinelandii*: removal of excess reductant by the cytochrome *d* pathway. Appl Environ Microbiol. 61: 3998–4003.

Moshiri. F., Crouse, B. R., Johnson, M. K. and Maier, R. J. 1995. The 'nitrogenase protective' FeSII protein of *Azotobacter vinelandii*: overexpression, characterization and crystallization. Biochemistry. 34: 12973–12982.

Poole, R. K. and Hill, S. 1997. Respiratory protection of nitrogenase activity in *Azotobacter vinelandi*i: roles of the terminal oxidases Bioscience Rep. 17: 303-317

Postgate, J. R. 1998. Nitrogen fixation, 3rd ed., Cambridge University Press, Cambridge

Rediers, H., Vanderleyden, J. and De Mot, R. 2000. *Azotobacter vinelandii*: a *Pseudomonas* in disguise? Microbiology. 150: 1117-1119.

Renn, S. C., Aubin-Horth, N., Hoffman, H. A. 2004. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. BMC Genomics 5: 42

Sabra, W., Zeng, A.-P., Lunsdorf, H. and Deckwer, W.-D. 2000. Effect of oxygen on formation and structure of *Azotobacter vinelandii* alginate and its role in protecting nitrogenase. Appl. Environ. Microbiol. 66: 4037-4044.

Strandberg, G. W. and Wilson, P. W. 1968. Formation of the nitrogen-fixing enzyme system in *Azotobacter vinelandii*. Can. J. Microbiol. 14(1): 25-31